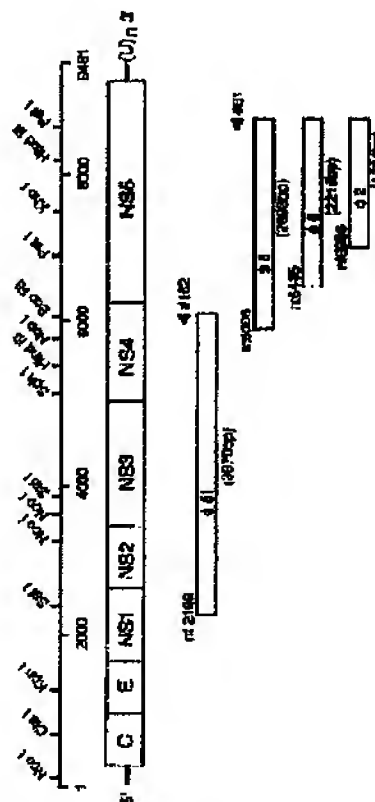


Patent number: JP6121689
Publication date: 1994-05-06
Inventor: OKAMOTO HIROAKI; NAKAMURA TETSUO
Applicant: NAKAMURA TETSUO
Classification:
- international: *A61B10/00; A61K39/29; A61K39/395; C12N15/51; C12P21/08; G01N33/53; G01N33/576; G01N33/577; C12P21/02; A61B10/00; A61K39/29; A61K39/395; C12N15/51; C12P21/08; G01N33/53; G01N33/576; G01N33/577; C12P21/02; (IPC1-7): A61B10/00; C12N15/51; A61K39/29; A61K39/395; C07K13/00; C12P21/02; C12P21/08; G01N33/53; G01N33/576; G01N33/577*
- european:
Application number: JP19910287402 19910809
Priority number(s): JP19910287402 19910809

Report a data error here

PURPOSE:To elucidate the whole genetic sequence of a non-A non-B (NANB) hepatitis virus HC-J6 strain and detect the NANB hepatitis virus with high sensitivity based thereon. **CONSTITUTION:**This is an invention of a NANB hepatitis virus gene and relates to a polynucleotide, a cDNA clone, a polypeptide, a polyclonal antibody and a monoclonal antibody related thereto and further a NANB hepatitis virus detection system using them. Thereby, this antibody is capable of detecting the NANB hepatitis virus with a high sensitivity and the related polynucleotide and cDNA clone are capable of producing the detection system.



Data supplied from the *esp@cenet* database - Worldwide

1

【特許請求の範囲】

【請求項1】配列番号1記載の塩基配列を有する非A非B型肝炎ウイルス遺伝子。

【請求項2】RNA 5'末端より340箇のヌクレオチドからなる非コード領域、これに続く構造蛋白質および非構造蛋白質をコードする領域を含む9099箇のヌクレオチド、ならびに3'末端108箇のUストレッチを含む150箇のヌクレオチドからなる、請求項第1項記載のウイルス遺伝子。

【請求項3】請求項第1項または第2項記載の遺伝子と部分的置換を有する非A非B型肝炎ウイルス変異株遺伝子。

【請求項4】配列番号2記載の塩基配列を有するポリヌクレオチドN-9589。

【請求項5】配列番号3記載の塩基配列を有するcDNAクローンJ6-φ81。

【請求項6】配列番号4記載の塩基配列を有するcDNAクローンJ6-φ8。

【請求項7】請求項第4項または第6項記載のヌクレオチドと部分的置換を有する非A非B型肝炎ウイルス変異株ヌクレオチド。

【請求項8】請求項第1項ないし第7項記載の遺伝子またはポリヌクレオチドにコードされるポリペプチド。

【請求項9】配列番号5記載のアミノ酸配列を有するポリペプチドP-3033。

【請求項10】請求項第1項ないし第7項記載の遺伝子、ポリヌクレオチド、cDNAクローンの全部または一部を宿主細胞に組み込み発現させて得たポリペプチド。

【請求項11】請求項第8項ないし第10項記載のポリペプチドに対するポリクローナル抗体またはモノクローナル抗体。

【請求項12】請求項第8項ないし第10項記載のポリペプチドを用いた非A非B型肝炎検出系。

【請求項13】請求項第11項記載の抗体を用いた非A非B型肝炎検出系。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、非A非B型肝炎ウイルス（以下「NANB型肝炎ウイルス」と略記する）遺伝子、これに関連するポリヌクレオチド、ポリペプチド、ならびに関連抗原、抗体検出系に関する。

【0002】

【従来の技術】ウイルスに由来する肝炎には、既に病原体が解明されその診断法や予防法が確立したものととしてA型肝炎、B型肝炎がある。これ以外は一括して、非A非B型肝炎（以下「NANB型肝炎」という）と呼ばれてきた。日本における輸血後肝炎は、B型肝炎の検出法が輸血血液のスクリーニングに導入された後は格段に減少したが、なお、年間推定28万例が、NANB型肝炎ウイ

2

ルスにより発症していると考えられてきた。

【0003】近年、NANB型肝炎ウイルス肝炎は、タイプ別にC型、D型、E型と命名され、その病原体についての研究が進展し、撲滅に向けての努力が世界的に進められるようになった。輸血後肝炎に関して、1988年アメリカのカイロン社がC型肝炎ウイルス（以下「HCV」と略記する）のRNAウイルスゲノムのクローン化に成功したと発表し、これを基にHCV抗体の測定系を開発した。現在、HCV抗体測定系は、輸血血液のスクリーニングや肝疾患患者の診断に、日本を始めとして各国で用いられている。このHCV抗体の測定系はNANB肝炎との関連性が確かに一部では認められる。しかし、キャリアや慢性肝炎の捕捉率は約70%にすぎず、また急性期の抗体検出ができない等重大な問題点が残されている。したがって、カイロン社の前記開発によってもNANB肝炎への対処は依然として解決されていない。NANB肝炎の経過は不良であり、水平感染によっても持続感染化（キャリア化）し、慢性肝炎に進展することが確認されている。また、肝硬変、肝がんへ進展するケースも多いと予想され、一日も早いウイルス本体の全貌解明が望まれていた。

【0004】

【発明が解決しようとする課題】上記カイロン社のHCV抗体検出キットでは診断できないNANB肝炎が多数存在することから、日本におけるNANB型肝炎ウイルスゲノムがカイロン社の解明したHCVゲノムとは必ずしも一致せず、サブタイプの違いがあると推定された。したがって、特異性、感度においてさらに優れたNANB肝炎診断薬、および有効なワクチンを開発するために、サブタイプ別にNANB型肝炎ウイルスの遺伝子レベル、アミノ酸レベルでの解析を完成することが望まれていた。本発明の目的は、NANB型肝炎ウイルスゲノムの塩基配列を、全体にわたって完全に解明し、これに基づいてアミノ酸レベルでの解析を行うこと、ならびにこれを用いてNANB型肝炎ウイルス関連抗原、抗体の検出系の確立を図ることである。

【0005】

【課題を解決するための手段】本発明者らは、NANB型肝炎ウイルスの治療、予防、検査に不可欠なポリヌクレオチドを見出すために鋭意研究をつづけ、その結果、ヒトおよびチンパンジーキャリアの血清よりNANB型肝炎ウイルスのRNAを単離し、非コード領域を含む遺伝子領域全体をカバーするcDNAをクローン化してその全部の塩基配列を決定し、さらにアミノ酸配列を決定した。その結果本発明者らは、NANB型肝炎ウイルスであるHCV-J6ゲノムの遺伝子全体の塩基配列を完全に解明することに成功した。かかる遺伝子及びアミノ酸レベルでの研究によって本発明は完成された。本発明のNANB型肝炎ウイルス（HCV-J6ゲノム）は、カイロン社のHCVゲノムとは別異のものである。

3

【0006】本発明は、配列番号1記載の塩基配列を有する非A非B型肝炎ウイルス遺伝子の発明であり、RNA 5'末端より340箇のヌクレオチドからなる非コード領域、これに続く構造蛋白質および非構造蛋白質をコードする領域を含む9099箇のヌクレオチド、ならびに3'末端108箇のUストッチを含む150箇のヌクレオチドからなる、配列番号1記載のウイルス遺伝子の発明であり、これらと、部分的置換を有する非A非B型肝炎ウイルス変異株遺伝子の発明である。また、配列番号2記載の塩基配列を有するポリヌクレオチドN-9589、配列番号3記載の塩基配列を有するcDNAクローンJ6-φ81、配列番号4記載の塩基配列を有するcDNAクローンJ6-φ8、これら各ヌクレオチドと部分的置換を有する非A非B型肝炎ウイルス変異株ヌクレオチドの発明である。さらに前記遺伝子またはポリヌクレオチドにコードされるポリペプチド、配列番号5記載のアミノ酸配列を有するポリペプチドP-3033、前記遺伝子、ポリヌクレオチド、cDNAクローンの全部または一部を宿主細胞に組み込み発現させて得たポリペプチド、前記各ポリペプチドに対するポリクローナル抗体またはモノクローナル抗体の発明である。さらに、前記各ポリペプチドを用いた非A非B型肝炎検出系、前記各抗体を用いた非A非B型肝炎検出系の発明である。

【0007】本発明のNANB型肝炎ウイルス遺伝子は次のようにして得て、その塩基配列を決定した。ヒトおよびチンパンジー血漿より得た検体、HC-J1、HC-J4、HC-J6を用いて本研究を進めた。HC-J1、HC-J6は日本人供血者由来で、いずれもHCV抗体陽性と判定された検体より得た。HC-J4はNANB肝炎の感染性を確認したチンパンジーから得た検体に由来するが、しかしこれはカイロン社の前記のHCV抗体では(一)であった。この3種の血漿検体よりRNAを調整し、これらによってRNAの塩基配列を調べた。このうちRNAの5'末端側の約2500、および3'末側の約1100の塩基配列については既に本年4月26日付特許出願(特願平3-191376号)ならびに6月20日付特許出願においてその一部を開示した。本発明者らは、さらに研究を進め、今回HC-J6について、新たに非構造蛋白質をコードする領域のcDNAクローンの構造も解明し、これによって、HC-J6ゲノムの完全長RNA9589塩基の配列の研究を完成させ、本発明を完成させるに至った。

【0008】HC-J6ゲノムは、実施例に示すように、5'末端より340箇のヌクレオチドよりなる5'非コード領域を有し、それに続いて構造蛋白質をコードする領域が続く、さらに非構造蛋白質をコードする領域を有する。他方、3'末端側は、非構造蛋白質をコードする領域に続いて108塩基のU-ストッチを含む150塩基の非コード領域を有する。5'末端側から341

4

番目のアデニンより始まるコード領域は9099塩基の一本の長いオープンリーディングフレームよりなり、これが3033アミノ酸をコードしていることが明らかとなった。HCV、すなわちC型肝炎ウイルスはその遺伝子構造からフラビウイルスに近縁であると考えられているが、本発明のNANB型肝炎ウイルス遺伝子のコード領域も図1に示す如く、C(コア)、E(エンベロープ)、NS-1(非構造蛋白-1)、NS-2(同-2)、NS-3(同-3)、NS-4(同-4)、NS-5(同-5)からなると考えられる。

【0009】本発明のHC-J6ゲノムは、ヨーロッパ特許公開第388,232号(カイロン社)に示されたHCVゲノムと比較すると、ヌクレオチド配列について全領域では67.9%の相同性を、またアミノ酸配列については全領域で72.3%の相同性を示すにすぎない。領域ごとに個別に相同性を調べると、5'末側非コード領域において塩基配列で94.4%、コア領域においてアミノ酸配列で90.1%であり、これらの領域では比較的高い相同性を示している。他方、エンベロープ部分より下流の領域ではE、NS-1、NS-2、NS-3、NS-4、NS-5のそれぞれのアミノ酸配列で60.4%、71.1%、57.8%、81.1%、73.1%、69.9%、3'末側非コード領域の塩基配列で29.6%といずれも低い相同性を示すにすぎず、結局、HC-J6株がカイロン社の発見したHCV株と大きく異なることが明らかとなった。

【0010】本発明者ら既に公表した(Japan. J. Exp. Med., Vol. 60, 3, 167-177)HC-J1(アメリカ型)、HC-J4(日本型)株とのアミノ酸配列の比較では、コア領域では90%以上の相同性を示すのに対して、エンベロープ領域ではそれぞれ60.9%、53.1%と低い相同性を示すにすぎなかった。本発明により、HC-J6株は、HC-J1株、HC-J4株のいずれとも異なる別箇のタイプのウイルスであることが明らかとなった。

【0011】

【作用】本発明の遺伝子、ポリヌクレオチド、cDNAクローンは、その全部または一部を大腸菌、枯草菌等の宿主細胞に組み込んでポリペプチドを発現させることができる。本発明のポリペプチドは、NANB型肝炎の検出系として用いることができ、さらに、これに対するポリクローナル抗体またはモノクローナル抗体を製造することができる。本発明のポリクローナル抗体またはモノクローナル抗体は、非A非B型肝炎検出系として用いることができる。

【0012】

【実施例】以下、本発明の実施例について述べるが、もとより本発明がこれらの実施例に限定されるものではない。

【0013】実施例1

50

5

(1) RNA抽出

日本人献血者の血漿から得た、HCV抗体陽性（オーソ・ダイアグノスティック・システムズ社製、オーソHCV Ab ELISAキット）と判定された検体HC-J1、HC-J6およびNANB肝炎の感染を確認したチンパンジーから得たHCV抗体（-）の検体HC-J4より、次のようにしてRNAを抽出した。血漿にトリス塩酸緩衝液（10mM、pH8.0）を加え、 6.8×10^3 rpmで1時間遠心した。得られたペレットに200mM NaCl、10mM EDTA、2% (w/v) ドデシル硫酸ナトリウム（SDS）、および1mg/mlのプロテナーゼKを含むトリス塩酸緩衝液（50mM、pH8.0）を加え、60℃で1時間加温し、エタノール沈澱を行い、RNAを得た。

【0014】(2) HC-J1 cDNA合成

HC-J1血漿より抽出したRNAを70℃で1分間加温し、これを鋳型として10ユニットの逆転写酵素（cDNA Synthesis System Plus、アマシャム社）およびオリゴヌクレオチドプライマー（20-mer）20pmolを加えて、42℃、15時間反応させてcDNAを得た。プライマーとしては、ヨーロッパ特許出願公開318,216号に示されたHCV塩基配列を参照して合成した#8（5'-GATGCTTGCGGAACGCAATCA-3'）を用いた。

【0015】(3) ポリメラーゼチェーンリアクション（PCR）によるcDNAの増幅

Gene Amp DNA増幅試薬キット（パーキンエルマー・シータス社）を用いたSaikiらの方法（Science 239, 487-491, 1988）により、DNAサーマイクラー（パーキンエルマー・シータス社）にて35サイクルのcDNA増幅を行った。

【0016】(4) cDNAライブラリーの構築による

nt 24~826

#32 (5'-ACTCCACCATAGATCACTCC-3')

#122 (5'-AGGTTCCCTGTTGCATAATT-3')

クローン: C9397、C9388、C9764

nt 732~1907

#50 (5'-GCCGACCTCATGGGGTACAT-3')

#128 (5'-TCGGTCGTGCCCACTACCAAC-3')

クローン: C9316、C9752、C9753

nt 1867~2591

#149 (5'-TCTGTGTGTGGCCCACTGTA-3')

#146 (5'-AGTAGCATCATCCACAAGCA-3')

クローン: C11621、C11624、C11655

さらに上流の5'末端側を特定するため、nt 246~265のアンチセンスプライマー#36（5'-AACACTACTCGGCTAGCAGT-3'）を用いてcDNAを合成したのち、ターミナルデオキシヌクレオチジルトランスフェラーゼによりcDNAの5'末端に

6

HC-J1、HC-J4の5'末端側塩基配列の決定
cDNAライブラリーを用いたHC-J1ゲノムの5'末端側の塩基配列解析は図2（HC-J1）図3（HC-J4）に示すように、cDNAをバクテリオファージλgt10に挿入して得たクローンの解析及びcDNAをPCRにて増幅して得られたクローンの解析結果の両者を併せて決定した。図2および図3はNANB肝炎ウイルスゲノムの5'末端を制限酵素切断部位とともに示し、使用するプライマーの位置も示す。図中、実線はバクテリオファージλgt10のライブラリーによるクローンで塩基配列を決定した範囲を、点線はPCRによるクローンで塩基配列を決定した範囲を示す。HC-J1のnt 454~2109の1656塩基はプライマー#8から得たcDNAをλgt10ファージベクター（アマシャム社）に挿入して得られたクローンφ41により決定した。つぎにこのシーケンスをもとに合成した、nt 824~843の新しいプライマー#25（5'-TCCCTGTTGCATAGTTACAG-3'）を用いて、HC-J4のcDNAライブラリーから順次4つのcDNAクローンφ60、φ61、φ66、φ75を得て、これらのコンセンサスシーケンスから上流のnt 18~843の塩基配列を決めた。

【0017】(5) HC-J6、5'末端側塩基配列の決定

HC-J6ゲノムの5'末端側の塩基配列は図4に示すように、cDNAをPCRにて増幅して得られたクローンの解析結果により決定した。HC-J6検体について、抽出したRNAから前記(2)の方法に従って、cDNAを抽出し、そのシーケンスを決定した。nt 24~2551はHC-J4の塩基配列を基にしたプライマーの組み合わせによるPCRにて得られた各クローンのコンセンサスシーケンスから決められた。

dA付加を行い、2段階のone-sided PCR増幅を行った。すなわち、1回目はオリゴdTプライマー（20-mer）とnt 188~207のアンチセンスプライマー#48（5'-GTTGATCCAAGAAAGGACC-3'）を用いて35サイクルのPCR

7

増幅を行い、2回目はそのPCR産物を鋳型にしてオリゴdTプライマー(20-mer)とnt140~160のアンチセンスプライマー#109(21-mer; 5'-ACCGGATCCGCAGACCACTAT-3')を用いて30サイクルのPCRを行った。得られたPCR産物をM13ファージベクターにサブクローニングした。完全長の5'末端配列を有すると考えられる13個の独立したクローン、C9577、C9579、C9581、C9584、C9587、C9590、C9591、C9595、C9606、C9609、C9615、C9616、C9619を得、それらのコンセンサスシーケンスからnt1~23の塩基配列を決めた。

(6) HC-J6、中央領域塩基配列の決定

HC-J6血漿100mlを原料に、(2)の方法に準じてλgt10バクテリオファージにcDNAライブラリーを構築した。プライマーとして、ヨーロッパ特許出公開318、216号に示されたHCV塩基配列を参照して合成したポリヌクレオチド#162および#81を用い、クローンの選択はブラークハイブリダイゼーションにより行った。図1に示すように、得られた4個のcDNAクローンφ2(nt6996~8700)、φ6(nt6485~8700)、φ8(nt6008~8700)、φ81(nt2199~6168)を解析し、得られたコンセンサスシーケンスからnt2552~8700の塩基配列を決定した。このうち、クローンφ81、φ8はそれぞれ配列番号3、4の塩基配列を有する。

【0018】(7) HC-J6ゲノムの3'末端側の塩基配列は図5に示すように、PCRにて増幅して得られたクローンの解析結果より決定した。HC-J6のnt8701~9241の配列については検体をプライマー#80(5'-GACACCCGCTGTTTGTGACTC-3')および#60(5'-GTTCTTACTGCCAGTTGAA-3')を用いたPCRにかけ、得られた938塩基のクローンC9760、C9234、C9761の塩基配列より得られるコンセンサスシーケンスから決定した。nt9242より下流の3'末端側塩基配列は以下の方法の従って決定した。すなわち、各検体より(1)記載の方法に従ってRNAを抽出し、ポリ(A)ポリメラーゼを用いてRNAの3'末端にポリ(A)を付加し、オリゴ(dT)20をプライマーとしてcDNA合成を行い、得られたcDNAをテンプレートとしてPCRに供した。PCRは第1段階としてセンスプライマーとして#97(5'-AGTCAGGGCGTCCCTCATCT-3')を、アンチセンスプライマーとしてオリゴ(dT)20を用いて行った。次に、得られたcDNAを第2段階PCRとして第1段階より下流に相当するセンスプライマー#90

8

(5'-GCCGTTTGCGGCCGATATCT-3')を、アンチセンスプライマーとしてオリゴ(dT)20を用いて行った。2段階のPCRで得られた増幅産物にT4DNAポリラーゼを作用させ両端を平滑にした後、T4ポリヌクレオチドキナーゼにて5'末端をリン酸化し、M13mp19ファージベクターのHincII部位にサブクローニングして塩基配列を決定した。こうして得られた19個のクローン、C10311、C10313、C10314、C10320、C10322、C10323、C10326、C10328、C10330、C10333、C10334、C10336、C10337、C10345、C10346、C10347、C10349、C10350、C10357から得られるコンセンサスシーケンスより3'末端側の塩基配列を決定した。以上の解析結果よりHC-J6ゲノムに相当するcDNA塩基配列を配列番号2記載のように決定し、またゲノムの完全長RNA塩基配列を配列番号1記載のように決定した。

【0019】(8) HC-J6、アミノ酸配列の決定

HC-J6ゲノムのRNA塩基配列をもとに、開始コドンATGより始まるコード領域のアミノ酸配列を配列番号5のように決定した。その結果、HC-J6は3033アミノ酸残基からなるポリペプチド前駆体をコードする1本の長いオープンリーディングフレームよりなることが明らかになった。

【0020】

【発明の効果】本発明は、非A非B型肝炎ウイルスHC-J6を単離し、そのウイルス遺伝子の全塩基配列を解明したことにもとづく発明であって、非A非B型肝炎ウイルスを高感度に検出することができ、また、これに必要なポリペプチド、ポリクローナル抗体、モノクローナル抗体を提供することができる。

【図面の簡単な説明】

【図1】は、本発明のNANB型肝炎ウイルス遺伝子のコード領域の構造。

【図2】～

【図4】は、NANB型肝炎ウイルスの5'末端側塩基配列決定の方法を示す図。

【図2】は、HC-J1株。

【図3】は、HC-J4株。

【図4】は、HC-J6株。

【図6】は、HC-J6株ゲノムの3'末端側塩基配列の決定方法。

【符号の説明】

【図1】において、Cはコア、Eはエンベロープ、NS-1は非構造蛋白質-1、NS-2は同-2、NS-3は同-3、NS-4は同-4、NS-5は同-5。

【配列表】

配列番号 : 1

配列の長さ : 9589

配列の型 : 核酸

鎖の数 : 一本鎖

トポロジー : 直鎖状

配列の種類 : genomic RNA

特徴を決定した方法 : E

配列

```

ACCCGCCCCU AAUAGGGGCG ACACUCCGCC AUGAACCACU CCCUGUGAG GAACUACUGU 60
CUUCACGCAG AAAGCGUCHA GCCAUGGCGU UAGUAUGAGU GUCGUACAGC CUCCAGGCCC 120
CCCCUCUCCG GGAGAGCCAU AGUGGUCUGC GGAACCGGUG AGUACACCGG AAUUGCCGGG 180
AAGACUGGGU CCUUUCUUGG AUAAACCCAC UCUAUGCCCG GUCAUUUGGG CGUGCCCCCG 240
CAAGACUGCU AGCCGAGUAG CGUUGGGUUG CGAAAGGCCU UGUGGUACUG CCUGAUAGGG 300
UGCUUGCGAG UGCCCCGGGA GGUCUCGUAG ACCGUGCACC AUGAGCACA AUCCUAAACC 360
UCAAGAAAA ACCAAAAGAA ACACCAACCG UGCCCCACAA GACGUUAAGU UUCCGGGCGG 420
CGCCAGAUJC GUUGGCGGAG UAUACUUGUU GCCGCGCAGG GGCCCCAGGU UGGGUGUGCG 480
CGCGACAAGG AAGACUUCGG AGCGGUCCCA GCCACGUGGA AGGCGCCAGC CCAUCCCUAA 540
GGAUCGGCGC UCCACUGGCA AAUCCUGGGG AAAACCAGGA UACCCUUGGC CCCUAAUACG 600
GAAUGAGGGA CUCGGCUGGG CAGGAUGGCU CCUGUCCCCC CGAGGUUCCC GUCCCUUUG 660
GGGCCCCAAU GACCCCCGGC AUAGGUCCCG CAACGUGGGU AAGGUCAUCG AUACCCUAC 720
GUGCGGCUUU GCCGACCUCA UGGGUACAU CCCUGUCGUA GCGCCCCGC UCGGCGGCGU 780
CGCCAGAGCU CUCGCGCAUG GCGUGAGAGU CCUGGAGGAC GGGGUUAUUU UUGCAACAGG 840
GAACUUACCC GGUUGCUECU UUUCUUCUU CUUGCUGGCC CUGCUGUCCU GCAUACCAC 900
CCCGUCUCC GCUGCCGAAG UGAAGAACAU CAGUACCGGC UACAUGGUGA CCAACGACUG 960
CACCAUGAU AGCAUUACCU GGCAACUCCA GGCUGCUGUC CUCCACGUCC CCGGGUGCGU 1020
CCCGUGCGAG AAAGUGGGGA AUACAUCUG GUGCUGGAUA CCGGUCUCAC CGAAUGUGGC 1080
CGUGCAGCAG CCCGGCGGCC UCACGCAGGG CUUACGGACG CACAUBACA UGGUUGUGAU 1140

```

GUCCGCCACG CUCUGCUCG CUCUUUACGU GGGGGACCUC UGCGGUGGGG UGAUGCUGC 1200
 AGCCAGAUU UUCAUUGUCU CCCCACAGCA CCACUGGUUU GUGCAAGACU GCAAUUGCUC 1260
 CAUCUACCCU GGUACCAUCA CUGGACACCG CAUGGCGUGG GACAUGAUGA UGAACUGGUC 1320
 GCCCACGGCU ACCAUGAUCC UGGCGUACGC GAUGCGCGUC CCCGAGGUA UCAUAGACAU 1380
 CAUUGGCGGG GCUCAUUGGG GCGUCAUGUU CGGCUUAGCC UACUUCUCUA UGCAGGGAGC 1440
 GUGGGCAAAA GUGGUUGUCA UUCUUUUGCU GGGCGCGGG GUGGACGCGC AAACCCAUAC 1500
 CGUUGGGGGU UCUACCGCGC AUAACGCCAG GACCCUCACC GGCAUGUUCU CCCUUGGUGC 1560
 CAGGCAGAAA AUCCAGCUCA UCAACACCAA UGGCAGUUGG CACAUAACC GCACCGCCCU 1620
 GAACUGCAAU GACUCUUUGC ACACCGGCUU CCUCGCGUCA CUGUUCUACA CCCACAGCUU 1680
 CAACUCGUCA GGAUGUCCCG AACGCAUGUC CGCCUGCCGC AGUAUCGAGG CCUUUCGGGU 1740
 GGAUGGGGGC GCCUUA CAU AUGAGACAA UGUACCAAU CCAGAGGAUA UGAGACCGUA 1800
 UUGCUGGCAC UACCCACCAA GACAGUGUGG UGUAGUCUCC GCGAGCUCUG UGUGUGGCC 1860
 AGUGUACUGU UUCACCCCCA GCCCAGUAGU AGUGGGUAGC ACCGAUAGAC UUGGAGCGCC 1920
 CACUACACG UGGGGGGAGA AUGAGACAGA UGUUUCUUA UUGAACAGCA CUCGACCACC 1980
 GCAGGGGUA UGGUUCGGCU GCACGUGGAU GAACUCCACU GGUACACCA AGACUUGCGG 2040
 CGCACACCC UGCCGCAUUA GAGCUGACUU CAUUGCCAGC AUGGACUUGU UGUGCCCCAC 2100
 GGACUGUUU AGGAAGCAUC CUGAUACCAC CUACAUCAAA UGUGGCUCUG GGGCCUGGCU 2160
 CAGCCAAGG UGCCUGAUCG ACUACCCCUA CAGGCUCUGG CAUUAACCCU GCACAGUUA 2220
 CUUACCAUC UUCAAAUUA GGAUGUAUGU GGGGGGGUC GAGCACAGG UCACGGCUGC 2280
 GUGCAAUUUC ACUCGUGGG AUCCUGGCAA CUUGGAGGAC AGAGACAGAA GUCAACUGUC 2340
 UCCUUUGCUG CACUCCACCA CGGAGUGGGC CAUUUUACCU UGCACUACU CGGACCUGCC 2400
 CGCCUUGUG ACUGGUCUUC UCCACCUCU CCAAAACAUC GUGGACGUGC AAUUCAGUA 2460
 UGGCCUAUA CCUGCUCUA CAAAUA CAU CGUCCGAUGG GAGUGGGUAG UACUCUUAU 2520
 CCUGCUCUA GCGGACGCA GGGUUUGGC CUGCUUAGG AUGCUCUUCU UGUUGGGCCA 2580
 GGGCAAGCA GCACUAGAGA AGUUGGUGU CUUGCACGU GCGAGCGCAG CUAGCUGCAA 2640
 UGGCUCCUA UACUUUGUA UCUUUUUGCU GGCUGCUUGG UACAUAAGG GUCGGGUAGU 2700
 CCCCUGGCU ACUUAUCCC UCACUGGCCU AUGGUCUUU GGGCUACUGC UCCUAGCAU 2760
 GGGCAACAG GCUAUUGCU AUGACGAUC UGUACAUGU CAGAUAGGAG CAGCUCUGU 2820
 GGUACUGAUC ACUCUCUUA CACUACCCC GGGUAUAAG ACCCUUCUA GCGGUUUCU 2880

GUGGUGGUG UGCUAUCUUC UGACCCUGGC GGAAGCUAUG GUCCAGGAGU GGGCACCACC 2940
 UAUGCAGGUG CGCGGUGGCC GUGAUGGGAU CAUAUGGGCC GUCCGCAUAU UCUGCCCGGG 3000
 UGUGGUGUUU GACAUAAACA AGUGGCUCUU GCGGUGCUU GGGCCUGCUU AUCUCCUAAA 3060
 AGGUGCUUUG ACGCGUGUGC CGUACUUCGU CAGGGCUCAC GCUCUACUAA GGAUGUGCAC 3120
 CAUGGUAAGG CAUCUCGCGG GGGUAGGUA CGUCCAGAUG GUGCUACUAG CCCUUGGCAG 3180
 GUGGACUGGC ACUUAACUUC AUGACCACCU CACCCCUAUG UCGGAUUGGG CUGCUAAUGG 3240
 CCUGCGGGAC UUGGCGGUGC CCGUGGAGCC UAUCAUCUUC AGUCCGAUGG AGAAAAAGU 3300
 CAUGGUCUGG GAGCGGAGA CAGCUGCUUG CGGGGAUAC UUAACCGAC UCCCGUGUC 3360
 CGCCGACUU GCGCGGAGG UCCUCCUUG CCCAGCUGAU GGUUAUACCU CCAAGGUG 3420
 GAGUCUUCUC GCCCCAUCA CUCUUAUGC CCAGCAGACA CGCGCCUUU UGGGACCAU 3480
 AGUGGUGAGC AUGACGGGC GCGACAAGC AGAACAGGCC GGGGAGAUUC AGGUCCUGUC 3540
 CACGGUACU CAGUCCUUC UCGGAACAAC CAUCUCGGG GUCUUAUGGA CUGUCUACCA 3600
 UGGAGCUGGC AACAGACUC UAGCGGCUC ACGGGUCCG GUACACAGA UGUACUCCAG 3660
 UGCUGAGGGG GACUAGUGG GGUGGCCAG CCCCCCGG ACCAAUCUU UGGAGCCUG 3720
 CAGGUGUGGA GCGGUCGAC UAUACUGGU CAGCGAAAC GCUGAUGUA UCCCGGUCG 3780
 AAGACGCGG GACAAGCGAG GAGCGUACU CUCGCGAGA CCUCUUCCA CCUUGAAGG 3840
 GUCCUCGGG GCGCGGUGC UCUGCCCAG AGGCACGCU GUCGGGUCU UCCGGGAGC 3900
 CGUGGUCUC CCGGGGUGG CCAAGUCCAU AGAUUUUUC CCCGUGAGA CACUUGACAU 3960
 CGUCACUCG UCCCCACCU UAAGUGACAA CAGCACACCA CCUGCUGUC CCAAACUUA 4020
 UCAGGUCGG UACUACAUG CCGGACUGG UAGUGGAAAG AGCACCAGG UCCUGUCG 4080
 GUUUGCGCU CAGGGGUA AAGUGCUAG GCUUAUCC UCGGUGGUG CCACCCUGG 4140
 GUUUGGGCG UACUUGUCA AGGCACUGG CAUCAAUCC AACAUUAGGA CUGGGGUCAG 4200
 GACUGGAGC ACCGGGCGC CCAUCACGUA CUCCACUAU GGCAAAUCC UCGCGAUGG 4260
 GGGCUGCGA GCGGCGCCU AUGACAUCU CAUAUGCGAU GAAUGCCAG CCGUGGACUC 4320
 UACCACCAU CUGGCAUCG GAACAGUCC CGAUCAAGCA GAGACAGCG GGUACAGCU 4380
 AACUGUACG GCUACGCUA CGCCCCCGG GUCAGUGACA ACCCCCCACC CCAACAUGA 4440
 GAGGUGGCC CUCGGCAGG AGGUGAGAU CCCUUCUUA GGGAGGGGA UUCCCUGUC 4500
 AUACAUAAG GAGGAAGAC ACUUGAUCU CUGCCACUA AAGAAAAAG UGACGAGCU 4560
 CGCGCGGCC CUUCGGGUA UGGGUUGAA CGCAGUGGA UACUACAGG GGUUGACGU 4620

CUCCGUAAUA CCAACUCAGG GAGACGUAGU GGUCGUCGCC ACCGACGCCC UCAUCACGGG 4680
 GUUUACUGGA GACUUUGACU CCGUGAUCGA CUGCAACGUA GCGGUCACUC AAGUUGUAGA 4740
 CUUCAGCUUG GACCCACAU UCACCAUAAC CAGACAGACU GUCCCUCAAG ACGCUGUCUC 4800
 ACGUAGCCAG CGCCGGGCC GCACGGGCAG GGAAGACUG GGUAUUUAUA GGUAUGUUUC 4860
 CACUGGUGAG CGAGCCUCAG GAAUGUUUGA CAGUGUAGUG CUCUGCGAGU GCUACGAUGC 4920
 AGGGGCCGCA UGGUAUGAGC UCACACCAGC GGAGACCACC GUCAGGCUCA GAGCAUAUUU 4980
 CAACACACCU GGUUGGCCUG UGUGCCAAGA CCAUCUUGAG UUUUGGGAGG CAGUUUUCAC 5040
 CGGCCUCACA CACAUAGAUG CCCACUCCU UCCCAACAA AAGCAAUCCG GGGAAAAUUU 5100
 CGCAUACUUA ACAGCCUACC AGGCUACAGU GUGCGCUAGG GCCAAAGCCC CCCCCCGUC 5160
 CUGGAGCUC AUGUGGAAGU GUUUGACUCG ACUCAAGCCC ACACUCGUGG GCCCCACACC 5220
 UCUCUGUAC CGCUUGGGCU CUGUUAACAA CGAGGUCACC CUCACGCATC CUGUGACGAA 5280
 AUACAUCCG ACCUGCAUGC AAGCCGACCU UGAGGUC AUG ACCAGCACGU GGGUCUAGC 5340
 UGGGGGGGUC UUGGCGGCCG UCGCCGCGUA CUGCCUGGCG ACCGGGUGUG UUGCAUCAU 5400
 CGGCCGCUUG CAGGUUAACC AGCGAGCCGU CGUUGCACCG GACAAGGAGG UCCUCUAUGA 5460
 GGCUUUGAU GAGUUGGAGG AAUGUGCCUC UAGAGCGGCU CUCAUUGAAG AGGGGCAGCG 5520
 GAUAGCCGAG AUGCUGAAGU CCAAGAUCCA AGGCUUAUUG CAGCAAGCUU CCAAACAAGC 5580
 UCAAGACAU CAACCCGCG UGCAGGCUUC UUGGCCAAG GUAGAGCAAU UCUGGGCAA 5640
 ACACAUUGG AACUUAUCA GCGGCAUUA AUACCUCCGA GGACUAUCAA CACUGCCAGG 5700
 GAACCCUGCU GUAGCUCCA UGAUGGCAU CAGUGCCGCC CUCACCAGUC CGUUGUAAC 5760
 UAGCACCACU AUCCUUCUA ACAUUUGGG GGGCUGGCUA GCAUCCCAA UUGCGCCUCC 5820
 CGCGGGGCU ACCGGCUUC UCGUAGUGG CCUGGUGGGG GCUGCCGUAG GCAGCAUAGG 5880
 CUUGGGUAG GUGCUGGUG ACAUCCUGG AGGUAUGGU GCGGGCAUUU CCGGGGCUU 5940
 CGUCGCAUUC AAGAUAUGU CUGGCGAGAA GCCUCCAU GAGGAUGUUG UCAACCUUCU 6000
 GCCUGGAUU CUGUCUCCG GUGCCUGGU GGUGGAGUC AUCUGCGCG CCAUCCUCC 6060
 CCGACACGUG GGACCGGGG AAGGCGUGU CCAAUUGAU AAUAGGCUA UUGCCUUGC 6120
 UCCAGAGGA AACCACGUG CCCCACCA CUACGUGAC GAGUCGGAUG CGUCGACGG 6180
 UGUGACCAA CUACUUGGU CCCUACCAU AACCAGCCUG CUCAGGAGC UCCACAACUG 6240
 GAUUACUGAA GACUGCCCA UCCAUAGCAG CGGUCGUGG CUCCGCGAUG UGUGGGAUUG 6300
 GGUUGCACC AUCCUACAG ACUUAAAAA CUGGUGACC UCCAAAUUG UCCCAAAGAU 6360

GCCUGGUCUC CCCUUUAUCU CUUGUCAAAA GGGGUACAAG GGCUGUGGG CUGGCACUGG 6420
 UAUCAUGACC ACACGGUGUC CUUGCGGGGC CAAUAUCUCU GGCAAUGUCC GCCUGGGCUC 6480
 CAUGAGAAUU ACGGGGCCCA AAACCUBCAU GAUAUUCUGG CAGGGGACCU UUCCCAUCA 6540
 UUGUACACG GAGGGCCAGU GCGUGCCGAA ACCCGCACCA AACUUUAAGA UCGCCAUCUG 6600
 GAGGGUGGCG GCCUCAGAGU ACGCGGAGGU GAGCGAGCAC GGUCAUACC ACUACAUAAC 6660
 AGGACUJACC ACUGAUAAU UGAAAGUUC UUGCCAACUA CCUUCUCCAG AGUUCUUUIC 6720
 CUGGGUGGAC GGAGUGCAGA UCCAUAGGUU UGCCCCAUA CCGAAGCCGU UUUUUGGGGA 6780
 UGAGGUCUCG UUCUGCGUUG GGCUUAUUC AUUUGUCGUC GGGUCUCAGC UCCCUUGCGA 6840
 UCCUGAACCU GACACAGAGC UAUGAGGUC CAUGCUACA GACCCAUCCC AUUACACGGC 6900
 GGAGACUGCA GCGCGGCGU UGGCAGGGG GUCACCCCG UCCGAGGCAA GCUCCUCAGC 6960
 GAGCCAGCUA UCGGACCAU CCGUGCGAGC CACCGCACC ACCCACGGCA AGGCCUAUGA 7020
 UGUGGACAUG GUGGAUGCCA ACCUGUUCAU GGGGGGCGAU GUGACCCGGA UAGAGUCUGA 7080
 GUCCAAGUG GUCGUCUGG ACUCUCUGA CCAAUGGUC GAAGAAAGGA GCGACCUUGA 7140
 GCCUUGGAUA CCAUGGAAU AUUGCUCCC CAAGAAGAGA UUGCCACCAG CCUUACCGGC 7200
 UUGGGCACGG CCUGAUUACA ACCCACCUCU UGUGGAUUG UGGAAGAGGC CAGAUUACCA 7260
 ACCGGCCACU GUUGCGGGCU GCGUCUCCC CCCCCUAAG AAAACCCGA CGCCUCCCC 7320
 AAGGAGAGC GGCAGUGG GUCUGAGUGA GAGCUCAUA GCAGUAGCC UACAACAGCU 7380
 GGCCAUCAAG UCCUUUGGC AGCCCCCCC AAGCGGCGAU UCAGGCCUUU CCACGGGGG 7440
 GCAGGAGGC GAUUCGGCA GUCGGAGGC CCGCAUGAG UUGGCCUUU CCGAGACAGG 7500
 UUGCAUCUC UCCAUGCCC CUCUGAGGG GAGCCUGGA GAUCCAGACU UGGAGCCUGA 7560
 GCAGGUAGG CUUCAACCUC CCCCCAGGG GGGGGUGGUA ACCCCCGGU CAGGCUCGGG 7620
 GUCUUGGUCU ACUUGCUCG AGGAGGAGC CUCGUCGUG UGCUGCUCA UGUCAUACUC 7680
 CUGGACCGGG GCUCUAAUA CUCUUGUAG CCGGAGAGG GAAAAGUUG CAAUUGGCC 7740
 CUUGAGCAAC UCCUGUUGC GAUAUCACA CAAGGUGUAC UGUACCACAU CAAAGAGCGC 7800
 CUCAUUAAG GCUAAAAGG UAACUUUGA UAGGAUGCAA GCUCUGAGC CUCAUUAUGA 7860
 CUCAGUCUG AAGGACAUU AGCUAGCGC CUCCAAGGUC ACCGCAAGG UUCUCACUUU 7920
 AGAGGAGGCC UGCAGUUA CUCACCCCA CUCUGCAAG UCCAAGUAG GGUUUGGGG 7980
 UAAGGAGGUC CGCAGCUUG CCGGAGAGC CGUUAACCAC AUCAAGUCCG UGUGGAAGGA 8040
 CCUCCUGGA GACACACAA CACCAAUUC UACAACCAUC AUGGCCAAA AUGAGGUGUU 8100

CUGCGUGGAC CCCACCAAGG GGGGUAAGAA AGCAGCUCGC CUUAUCGUUU ACCCUGACCU 8160
 CGGCGUCAGG GUCUGCGAGA AAAUGGCCCCU UUAUGAUUUC ACACAAAAGC UUCUCAGGC 8220
 GGUGAUGGGG GCUCUUAUG GAUUCAGUA CUCCCCCGCU CAGCGGGUGG AGUUUCUCUU 8280
 GAAGGCAUGG GCGGAAAAGA AAGACCCUUAU GGGUUUUUCG UAUGAUCCC GAUGCUUUGA 8340
 CUCAACCGUC ACUGAGAGAG ACAUCAGGAC UGAGGAGUCC AUUAUUCGGG CUUGUUCUU 8400
 GCGCGAGGAG GCGCACACUG CCAUACACUC ACUGACUGAG AGACUUUACG UGGGAGGGCC 8460
 CAUGUUAAC AGCAAGGGCC AGACCUGCGG GUACAGGCGU UGCCGCGCCA GCGGGGUGCU 8520
 UACCACUAGC AUGGGGAACA CCAUCACAUG CUAUGUGAAA GCCUWAGCGG CCUGUAAGGC 8580
 UGCAGGGAUA AUUGCGCCCA CAAUGCUGGU AUGCGGCGAU GACUUGGUUG UCAUCUCAGA 8640
 GAGCGAGGGG ACCGAGGAGG ACGAGCGGAA CCUGAGAGCC UUCACGGAGG CUAUGACCAG 8700
 GUUUCUGGC CCUCCUGGUG ACCCCCCCAG ACCGGAUAU GACCUGGAGC UGAUAACAUC 8760
 UUGCUCCUCA AAUGUGUCUG UGGCGUUGGG CCCACAAGGC CGCCGCAGAU ACUACCUGAC 8820
 CAGAGACCCU ACCACUCCAA UCGCCGGGGC UGCCU8GGAA ACAGUUAGAC ACUCCCCUGU 8880
 CAAUUAUGG CUAGGAAACA UCAUCCAGUA CGCCCCAACC AUUUGGGCUC GCAUGGUCCU 8940
 GAUGACACAC UUCUUCUCCA UUCUCAUGGC CCAAGAUACU CUGGACCAGA ACCUCAACUU 9000
 UGAGUUGUAC GGAGCGGUGU ACUCCGUGAG UCCCUUGGAC CUCCCAGCCA UAAUUGAAG 9060
 GUUACACGGG CUUGACGCUU UCUCUCUGCA CACAUACAU CCCCACGAAC UGACACGGGU 9120
 GGCUCAGCC CUCAGAAAAC UUGGGCGGCC ACCCCUCAGA GCGUGGAAGA GCGGGGCACG 9180
 UGCAGUCAGG GCGUCCUCA UCUCGCGUGG GGGGAGAGCG GCCGUUUGCG GCCGAUUCU 9240
 CUUCAACUGG GCGGUGAAGA CCAAGCUCAA ACUACUCCA UUGCCGGAAG CGCGCCUCCU 9300
 GGAUUUAUC AGCUGGUUCA CUGUCGGCGC CGGCGGGGCG GACAUUUAUC ACAGCGUGUC 9360
 GCGUGCCCGA CCCGCUUUAU UACUCCUUGG CCUACUCCUA CUUUUGUAG GGGUAGGCCU 9420
 UUUCCUACUC CCGCUCGCU AGAGCGGCAC ACAUUGCUA CACUCCAUG CUAACUGUCC 9480
 CUUUUUUUUU UUUUUUUUUU UUUUUUUUUU UUUUUUUUUU UUUUUUUUUU UUUUUUUUUU 9540
 UUUUUUUUUU UUUUUUUUUU UUUUUUUUUU UUUUUUUUUU UUUUUUUUUU 9589

配列番号 : 2

配列の長さ : 9589

配列の型 : 核酸

鎖の数 : 一本鎖

トポロジー : 直鎖状

配列の種類 : cDNA to genomic RNA

特徴を決定した方法 : E

配列

```

ACCCGCCCT AATAGGGCG ACACTCCGC ATGAACCACT CCCCTGTGAG GAACTACTGT   60
CTTCACGCAG AAAGCGTCTA GCCATGGCGT TAGTATGAGT GTCGTACAGC CTCACGGGCC   120
CCCCCTCCC GAGAGGCCAT AGTGGTCTGC GGAACCGGTG AGTACACCGG AATTGCCGGG   180
AAGACTGGGT CCTTCTCTGG ATAAACCCAC TCTATGCCCG GTCATTGGGG CGTGGCCCCG   240
CAAGACTGCT AGCCGAGTAG CGTTGGGTTG CGAAAGGCCT TGTGGTACTG CCTGATAGGG   300
TGCTTGGGAG TGCCCCGGGA GGTCTCGTAG ACCGTGCACC ATGAGCAGAA ATCCTAAACC   360
TCAAAGAAA ACCAAAAGAA ACACCAACCG TCGCCACAA GACGTTAAGT TTCCGGGGGG   420
CGGCCAGATC GTTGGCGGAG TATACTTGTT GCCCGGCAGG GCGCCGAGGT TGGGTGTGCG   480
CGCGACAAGG AAGACTTCGG AGCGGTCCCA GCCACGTGGA AGGCGCCAGC CCATCCCTAA   540
GGATCGGGCG TCCACTGGCA AATCCTGGGG AAAACCAAGG TACCCCTGGC CCCTATACGG   600
GAATGAGGGA CTCGGCTGGG CAGGATGGGT CCTGTCCCC CGAGGTTCCC GTCCCTCTTG   660
GGGCCCAAT GACCCCGGGC ATAGGTCCCG CAACGTGGGT AAGGTCATCG ATACCCTAAC   720
GTGCGGCTTT GCCGACCTCA TGGGGTACAT CCCTGTGTA GCGCCCCCGC TCGGCGGCGT   780
CGCCAGAGCT CTCGCGCATG GCGTGAGAGT CCTGGAGGAC GGGGTTAATT TTGCAACAGG   840
GAACCTACCC GGTGCTCCT TTTCTATCTT CTGCTGBCD CTGCTGCTCT GCATCACCAC   900
CCCGGTCTCC GCTGCCGAAG TGAAGAACAT CAGTACCGG TACATGGTGA CCAACGACTG   960
CACCAATGAT AGCATTACCT GSCAACTCCA GGCTGCTGTC CTCCACGTCC CCGGGTGCST 1020
CCCGTGCGAG AAAGTGGGGA ATACATCTCG GTGCTGGATA CCGGTCTCAC CGAATGTGGC 1080
CGTGACAGCAG CCCGGCGCCC TCACGCAGGG CTACGGAGG CACATTGACA TGGTTGTGAT 1140
GTCCGCCAGG CTCTGCTCG CTCTTTACGT GGGGGACCTC TGGGGTGGGG TGATGCTTGC 1200

```

AGCCAGATG TTCATTGTCT CGCCACAGCA CCACTGGTTT GTGCAAGACT GCAATTGCTC 1260
 CATCTACCCT GGTACCATCA CTGGACACCG CATGGCGTGG GACATGATGA TGAAGTGGTC 1320
 GCCACGGCT ACCATGATCC TGGCGTACGC GATGCGCGTC CCCGAGGTCA TCATAGACAT 1380
 CATTGGCGGG GCTCATTGGG GCGTCATGTT CCGCTTAGCC TACTTCTCTA TGCAGGGAGC 1440
 GTGGGCAAAA GTCGTTGTCA TTCTTTTGCT GCGCGCGGG GTGGACGGC AAACCCATAC 1500
 CGTTGGGGT TCTACCGCG ATAACGCCAG GACCCTCACC GGCATGTTCT CCCTTGGTGC 1560
 CAGGCAGAAA ATCCAGCTCA TCAACACCAA TGGCAGTTGG CACATCAACC GCACCGCCCT 1620
 GAAGTGAAT GACTCTTTGC ACACCGGCTT CCTCGCGTCA CTGTTCTACA CCCACAGCTT 1680
 CAACTCGTCA GGATGTCCCG AACGCATGTC CGCCTGCCGC AGTATCGAGG CCTTTCGGGT 1740
 GGGATGGGGC GCCTTACAAT ATGAGGACAA TGTACCAAT CCAGAGGATA TGAGACCGTA 1800
 TTGCTGGCAC TACCCACCAA GACAGTGTGG TGTAGTCTCC GCGAGCTCTG TGTGTGGCCC 1860
 AGTGTACTGT TTCACCCCA GCCCAGTAGT AGTGGGTACG ACCGATAGAC TTGGAGCGCC 1920
 CACTTACAGG TGGGGGAGA ATGAGACAGA TGTCTTCTA TTGAACAGCA CTCGACCACC 1980
 GCAGGGGTCA TGGTTCGGCT GCACGTGGAT GAAGTCCACT GGCACACCA AGACTTGGCG 2040
 CGCACCACC TCCCGCATTA GAGCTGACTT CAATGCCAGC ATGGACTTGT TGTGCCCCAC 2100
 GGACTGTTT AGGAAGCATC CTGATACCAC CTACATCAA TGTGGCTCTG GGCCTGGCT 2160
 CAGGCCAAGG TGCTGTATCG ACTACCCCTA CAGGCTCTGG CATTACCCCT GCACAGTTAA 2220
 CTATACCATC TTCAAATAA GGATGTATGT GGGGGGGGTC GAGCACAGG TCACGGCTGC 2280
 GTGCAATTTC ACTCGTGGG ATCGTTGCAA CTTGGAGGAC AGAGACAGAA GTCAACTGTC 2340
 TCGTTTGTG CACTCCACCA CGGAGTGGG CATTYTACCT TGCACTTACT CGGACCTGCC 2400
 CGCCTTGTG ACTGGTCTTC TCCACCTCCA CAAAACATC GTGGACGTGC AATTCATGTA 2460
 TGGCTATCA CCTGCTCTCA CAAATACAT CGTCCGATGG GAGTGGGTAG TACTCTTATT 2520
 CCTGCTCTTA GCGGACGCC GGGTTTGCGC CTGCTTATGG ATGCTCATCT TGTGGGCCA 2580
 GCGCGAAGCA GCACTAGAGA AGTTGGTCGT CTTGCACGCT GCGAGCGCAG CTAGCTGCAA 2640
 TGGCTTCTA TACTTTGTCA TCTTTTCGT GGCTGCTTGG TACATCAAGG GTCGGGTAGT 2700
 CCCCTTGGCT ACTTATGCC TCACTGCCCT ATGGTCCTT GGCCTACTGC TCCTAGCATT 2760
 GCGCCAACAG GTTATGCTT ATGACGCATC TGTACATGGT CAGATAGGAG CAGCTCTGTT 2820
 GGTACTGATC ACTCTCTTA CACTACCCC CGGGTATAAG ACCCTTCTCA GCGGTTTCT 2880
 GTGGTGGTTG TGTATCTTC TGACCCTGGC GGAAGCTATG GTCCAGGAGT GGGCACCACC 2940

TATGCAGGTG CGCGGTGCCC GTGATGGGAT CATATGGGCC GTCCCATAT TCTGCCCGGG 3000
 TGTGGTGTIT GACATAACCA AGTGGCTCTT GCGGTGCTT GGGCTGCTT ATCTCTAAA 3060
 AGGTGCTTTG ACGCGTGTGC CGTACTTCGT CAGGGCTCAC GCTCTACTAA GGATGTGCAC 3120
 CATGGTAAGG CATCTC8CGG GGGGTAGGTA CGTCCAGATG GTGCTACTAG CCCTTG6CAG 3180
 GTGGACTGGC ACTTACATCT ATGACCACCT CACCCCTATG TCGGATTGGG CTGCTAATGG 3240
 CCTGCGGGAC TTGGCGGTGG CCGTGGAGCC TATCATCTTC AGTCCGATGG AGAAAAAGT 3300
 CATCGTCTGG GGAGCGGAGA CAGCTGCTTG CGGGGATATC TTACACGGAC TTCCCGTGTG 3360
 CCCCCGACTT GGCCGGGAGG TCCTCCTTGG CCCAGCTGAT GGTATACCT CCAAGGGGTG 3420
 GAGTCTTCTC GCCCCATCA CTGCTTATGC CCAGCAGACA CGCGGCTTT TGGGCACCAT 3480
 AGTGGTGAAG ATGACGGGGC GCGACAAGAC AGAACAGGCC GGGGAGATTC AGGTCTGTG 3540
 CACGGTCACT CAGTCTTCC TCGGAACAAC CAICTGGGG GTCTTATGGA CTGTCTACCA 3600
 TGGAGCTGGC AACAAGACTC TAGCCGGCTC ACGGGGTCCG GTCACACAGA TGTACTCCAG 3660
 TGCTGAGGGG GACTTAGTGG GGTGGCCAG CCCCCCGGG ACCAAATCTT TGGAGCCGTG 3720
 CACGTGTGGA GCGGTGACC TATACCTGT CACCGGAAAC GCTGATGTCA TCCGGCTCG 3780
 AAGACCGGG GACAAGCGAG GAGCGTACT CTCCCCGAGA CCTCTTTCCA CCTTGAAGGG 3840
 GTCCTCGGGG GGGCGGTGC TCTGCCCCAG AGGCCACGT GTCGGGTCT TCGGGCAGC 3900
 CGTGTCTCC CGGGCGTGG CCAAGTCCAT AGATTTTATC CCCGTTGAGA CACTTGACAT 3960
 CGTCACTCGG TCCCCACCT TTAGTGACAA CAGCACACCA CCGCTGTGC CCCAACTTA 4020
 TCAGGTGGG TACTTACATG CCGGACTGG TAGTGAAG AGCACCAAG TCCCTGTGC 4080
 GTATGCGCT CAGGGGTACA AAGTGCTAGT GCTTAATCCC TCGGTGGCTG CCACCTGGG 4140
 GTTGGGGCG TACTGTCCA AGGCACATGG CATCAATCCC AACATTAGGA CTGGGTCAG 4200
 GACTGTGAG ACCGGGGCGC CCATCACGTA CTCCACATAT GGCAAAITCC TCGCCGATGG 4260
 GGGCTGCGCA GCGGGCGCT ATGACATCAT CATATGCGAT GAATGCCATG CCGTGGACTC 4320
 TACCACCATI CTCGGCATCG GAACAGTCT CGATCAAGCA GAGACAGCCG GGTCAAGCT 4380
 AACTGTACTG GCTACGGTA CCCCCCGG GTCAGTGACA ACCCCCCACC CCAACATAGA 4440
 GGAGGTGGC CTCGGGCAAG AGGTGAGAT CCCCTTCTAT GGGAGGGGCA TTCCCTGTG 4500
 ATACATCAAG GGAGGAAGAC ACTGATCTT CTGCCACTCA AAGAAAAAGT GTGACGAGCT 4560
 CCGGGCGGC CTTCGGGGA TGGCTTGAA CGCAGTGGCA TACTACAGAG GGCTGGAGCT 4620
 CTCGTAATA CCAACTCAGG GAGACGTAGT GTCGTGCGC ACCGACGCC TCATGACGGG 4680

GTTTACTGGA GACTTTGACT CCGTGATCGA CTGCAACGTA GCGGTCACTC AAGTTGTAGA 4740
 CTTCAGCTTG GACCCACAT TCACCATAAC CACACAGACT GTCCCTCAAG ACGCTGTCTC 4800
 ACGTAGCCAG CCGCGGGGCC GCACGGGCAG GGAAGACTG GGTATTTATA GGTATGTTTC 4860
 CACTGGTGAG CGAGCCTCAG GAATGTTTGA CAGTGTAGTG CTCTGCCAGT GCTACGATGC 4920
 AGGGGCCGCA TGGTATGAGC TCACACCAGC GGAGAGCACC GTCAGGCTCA GAGCATATTT 4980
 CAACACACCT GGTTCGCTG TGTGCCAAGA CCATCTTGAG TTTTGGGAGG CAGTTTTCAC 5040
 CCGCCTCACA CACATAGATG CCCACTTCCT TTCCCAAACA AAGCAATCGG GGGAAAATTT 5100
 CGCATACTTA ACAGCCTACC AGGCTACAGT GTGCGTAGG GCCAAAGCCC CCCCCCGTC 5160
 CTGGGACGTC ATGTGGAAGT GTTTGACTCG ACTCAAGCCC AACTCGTGG GCGCCACACC 5220
 TCCTCTGAC CGCTTGGGCT CTGTTACCA CGAGGTCACC CTCACGCATC CTGTGACGAA 5280
 ATACATCGCC ACCTGCATGC AAGCCGACCT TGAGGTCATG ACCAGCACGT GGGTCTTAGC 5340
 TGGGGGGGTC TTGGCGGCCG TCGCGCGTA CTGCTGGCG ACCGGGTGTG TTTGCATCAT 5400
 CGGCGCGTTC CAGGTTAACC AGCGAGCCGT CGTTGCACCG GACAAGGAGG TCCTCTATGA 5460
 GGCTTTTGAT GAGATGGAGG AATGTCCCTC TAGAGCGGCT CTCATTGAAG AGGGGCAGCG 5520
 GATAGCCGAG ATGCTGAAGT CCAAGATCCA AGGCTTATTC CAGCAAGCTT CCAAACAAGC 5580
 TCAAGACATA CAACCCGCTG TGCAGGCTTC TTGGCCCAAG GTAGAGCAAT TCTGGGCCAA 5640
 ACACATGTGG AACTTCATCA GCGGCATTCA ATACCTCGCA GGACTATCAA CACTGCCAGG 5700
 GAACCTGCT GTAGCTTCCA TGATGGCATT CAGTGCCGCC CTCACCAGTC CGTTGTCAAC 5760
 TAGCACCACCT ATCCTTCTCA ACATTTTGGG GGGCTGGCTA GCATCCCAA TTGCGCCTCC 5820
 CCGCGGGGCT ACCGGCTTCG TCGTCAGTGG CCTGGTGGGG GCTGCCGTAG GCAGCATAGG 5880
 CTTGGGTAA GGTCTGGTGG ACATCCTGGC AGGTATGGT GCGGGCATT CCGGGGCTCT 5940
 CGTCGCATTC AAGATCATGT CTGGCGAGAA GCGCTCCATG GAGGATGTTG TCAACCTGCT 6000
 GCCTGGAATT CTGTCTCCGG GTGCCCTGCT GGTGGGAGTC ATCTGCCCGG CCATCCTGCG 6060
 CCGACACGTG GGACCGGGGG AAGCGCTGT CCAATGGATG AATAGGCTCA TTGCCTTTGC 6120
 TTCCAGAGGA AACCACGTG CCCCCACCCA CTACGTGACG GAGTCGGATG CGTCGCAGCG 6180
 TGTGACCCAA CTACTTGGCT CCCTTACCAT AACCAAGCTG CTCAGGAGAC TCCACAAC TG 6240
 GATTACTGAA GACTGCCCCA TCCCATGCAG CGGCTCGTGG CTCCGGGATG TGTGGGATTG 6300
 GGTTCGACC ATCCTAACAG ACTTTAAAA CTGGCTGACC TCCAAATTGT TCCCAAAGAT 6360
 GCCTGGTCTC CCCTTATCT CTTGTCAAAA GGGGTACAAG GCGGTGTGG CTGGCACTGG 6420

TATCATGACC ACACGGTGTG CTGGCGGCGC CAATATCTCT GGCAATGTCC GCCTGGGCTC 6480
 CATGAGAATT ACGGGGCCCA AAACCTGCAT GAATATCTGG CAGGGGACCT TTCCCATCAA 6540
 TTGTTACACG GAGGGCCAGT GCGTGCCGAA ACCCGCACCA AACTTTAAGA TCGCCATCTG 6600
 GAGGGTGCGG GCCTCAGAGT ACCGGGAGGT GACGCAGCAC GGGTCATACC ACTACATAAC 6660
 AGGACTTACC ACTGATAACT TGAAGTTCC TTGCCAACTA CCTTCTCCAG AGTTCCTTTC 6720
 CTGGGTGGAC GGAGTGCAGA TCCATAGGTT TGCCCCATA CCGAAGCCGT TTTTCGGGA 6780
 TGAGGTCTCG TTCTGCGTTG GGCTTAATTC ATTTGTCGTC GGGTCTCAGC TCCCTTGCGA 6840
 TCCTGAACCT GACACAGACG TATTGACGTC CATGCTAACA GACCCATCCC ATATCAGGCG 6900
 GGAGACTGCA GCGCGGCGTT TGGCAGGGG GTCACCCCG TCCGAGGCAA GCTCCTCAGC 6960
 GAGCCAGCTA TCGGCACCAT CGCTGCGAGC CACCTGCACC ACCCAGGGA AGGCCTATGA 7020
 TGTGGACATG GTGGATGCCA ACCTGTTCAT GGGGGGCGAT GTGACCCGGA TAGAGTCTGA 7080
 GTCCAAAGTG GTCGTTCTGG ACTCTCTCGA CCCAATGGTC GAAGAAAGGA GCGACCTTGA 7140
 GCCTTCGATA CCATCGGAAT ATATGCTCCC CAAGAAGAGA TTCCACCAG CCTTACCGGC 7200
 TTGGGCACGG CCTGATTACA ACCCACCCTG TGTGAATCG TGAAGAGGC CAGATTACCA 7260
 ACCGGCCACT GTTGGGGCT GCGCTCTCCC CCCCCCTAAG AAAACCCCGA CGCCTCCCC 7320
 AAGGAGACGC CGGACAGTGG GTCTGAGTGA GAGTCCATA GCAGATGCCC TACAACAGCT 7380
 GGCATCAAG TCCTTTGGCC AGCCCCCCC AAGCGCGAT TCAGGCCCTT CCACGGGGGC 7440
 GGACGCAGCC GATTCCGGCA GTCGACGCC CCCCAGTAG TTGGCCCTTT CGGAGACAGG 7500
 TTCCATCTCC TCCATGCCCC CTCTCGAGG GGAGCCTGGA GATCCAGACT TGGAGCCTGA 7560
 GCAGGTAGAG CTTCAACCTC CCCCCAGGG GGGGGTGGTA ACCCCCGGCT CAGGCTCGGG 7620
 GTCTTGGTCT ACTTGCTCCG AGGAGGACGA CTCGTCGTG TGCTGCTCCA TGTCATCTC 7680
 CTGGACCGGG GCTCTAATAA CTCCTTGTAG CCCCAGAGG GAAAAGTTGC CAATTGGGCC 7740
 CTTGAGCAAC TCCCTGTTGC GATATCACA CAAGGTGTAC TGTACCACAT CAAAGAGCGC 7800
 CTCATTAAGG GCTAAAAAG TAACCTTTGA TAGGATGCAA GCGCTCGACG CTCATTATGA 7860
 CTCAGTCTTG AAGGACATTA AGCTAGCGGC CTCCAAGGTC ACCGCAAGGC TTCTCACTTT 7920
 AGAGGAGGCC TGCCAGTTAA CTCACCCCA CTCTGCAAGA TCCAAGTATG GGTTCGGGC 7980
 TAAGGAGGTC CGCAGCTTGT CCGGAGAGC CGTTAACCAC ATCAAGTCCG TGTGAAGGA 8040
 CCTCCTGGAA GACACACAAA CACCAATTCC TACAACCATC ATGGCCAAAA ATGAGGTGTT 8100
 CTGCGTGGAC CCCACCAAGG GGGGTAAGAA AGCAGCTCGC CTTATCGTTT ACCCTGACCT 8160

CGGCGTCAGG GTCTGCGAGA AAATGGCCCT TTATGATATC ACACAAAAGC TTCTCAGGC 8220
 GGTGATGGGG GCTTCTTATG GATTCCAGTA CTCCCCGCT CAGCGGGTGG AGTTTCTCTT 8280
 GAAGGCATGG GCGGAAAAGA AAGACCCCTAT GGGTTTTTCG TATGATACCC GATGCTTTGA 8340
 CTCAACCGTC ACTGAGAGAG ACATCAGGAC TGAGGAGTCC ATATATCGGG CTGTTCCTT 8400
 GCGCGAGGAG GCCCAGACTG CCATACACTC ACTGACTGAG AGACTTTACG TGGGAGGGCC 8460
 CATGTTCAAC AGCAAGGGCC AGACCTGCGG GTACAGGCGT TCCCGCGCCA GCGGGGTGCT 8520
 TACCACTAGC ATGGGGAACA CCATCACATG CTATGTGAAA GCCTTAGCGG CCTGTAAAGC 8580
 TGCAGGGATA ATTGCGCCCA CAATGCTGGT ATGCGGCGAT GACTTGCTTG TCATCTCAGA 8640
 GAGCCAGGGG ACCGAGGAGG ACGAGCGGAA CCTGAGAGCC TTCACGGAGG CTATGACCAG 8700
 GTATTCTGCC CCTCCTGGTG ACCCCCCAG ACCGGAATAT GACCTGGAGC TGATAACATC 8760
 TTGCTCCTCA AATGTGTCTG TGGCGTTGGG CCCACAAGGC CGCCGCAGAT ACTACCTGAC 8820
 CAGAGACCCT ACCACTCCAA TCGCCCGGGC TGCCCTGGAA ACAGTTAGAC ACTCCCTGT 8880
 CAATTCATGG CTAGGAAACA TCATCCAGTA CGCCCCAACC ATATGGGCTC GCATGGTCCT 8940
 GATGACACAC TTCTTCTCCA TTCTCATGGC CCAAGATACT CTGGACCAGA ACCTCAACTT 9000
 TGAGATGTAC GGAGCGGTGT ACTCCGTGAG TCCCTTGGAC CTCCAGCCA TAATTGAAAG 9060
 GTTACACGGG CTGACGCTT TCTCTCTGCA CACATACACT CCCCAGGAAC TGACACGGGT 9120
 GGCTTCAGCC CTCAGAAAC TTGGGGCGCC ACCCCTCAGA GCGTGGAAGA GCCGGGCACC 9180
 TGCAGTCAGG GCGTCCCTCA TCTCCGTTGG GGGGAGAGCG GCCGTTTGGC GCCGATATCT 9240
 CTTCAACTGG GCGGTGAAGA CCAAGCTCAA ACTCACTCCA TTGCCGAAG CGCGCTCCT 9300
 GGATTTATCC AGCTGTTCA CTGTGGGCGC CGCGGGGGC GACATTTATC ACAGCGTGC 9360
 GCGTGCCCGA CCCCGCTTAT TACTCCTTGG CCTACTCCTA CTTTTGTAG GGTAGGCCT 9420
 TTTCTACTC CCCGCTCGGT AGAGCGGCAC ACATTAGCTA CACTCCATAG CTAATGTCC 9480
 CTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT 9540
 TTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT 9589

配列番号：3

配列の長さ：3970

配列の型：核酸

鎖の数：一本鎖

トポロジー：直鎖状

配列の種類：cDNA to genomic RNA

特徴を決定した方法：E

配列

```

GECATTACCC CTGCACAGTT AACTATACCA TCTTCAAAAT AAGGATGTAT GTGGGGGGGG 60
TCGAGCACAG GCTCAGGGET GCGTGCAATT TCACTCGTGG GGATCGTTGC AACITGGAGG 120
ACAGAGACAG AAGTCAACTG TCTCCTTTGC TGCACTCCAC CACGGAGTGG GCCATTTTAC 180
CTTGCACTTA CTCGACCTG CCCGCCITGT CGACTGGTCT TCTCCACCTC CACCAAACA 240
TGGTGGACGT GCAATTCATG TATGGCTAT CACCTGCTCT CACAAATAC ATCGTCCGAT 300
GGGAGTGGGT AGTACTCTTA TTCCTGCTCT TAGCGGACGC CAGGGTTTGC GCCTGCTTAT 360
GGATGCTCAT CTGTGTGGG CAGGCCGAAG CAGCACTAGA GAAGTTGGTC GTCTTGACG 420
CTCGGAGCGC AGCTAGCTGC AATGGCTTCC TATACITTTG CATCTTTTTC GTGGCTGCTT 480
GGTACATCAA GGGTCGGGTA GTCCCTTGG CTACTIONTC CCTCACTGGC CTATGGTCCT 540
TTGGCCTACT GCTCCTAGCA TTGCCCAAC AGGCTTATGC TTATGACGCA TCTGTACATG 600
GTCAGATAGG AGCAGCTCTG TTGGTACTGA TCACTCTCTT TACACTCACC CCCGGGTATA 660
AGACCCITCT CAGCCGGTIT CTGTGGTGGT TGTGCTATCT TCTGACCCTG GCGGAAGCTA 720
TGGTCCAGGA GTGGGCACCA CCTATGCAAG TCGCGGTGG CCGTGATGG ATCATATGG 780
CCGTCCCAT ATTCTGCCG GGTGTGGTGT TTACATAAC CAAGTGGCTC TTGGCGGTGC 840
TTGGGCCTGC TTATCTCTA AAAGGTGCTT TGACGGGTGT GCGGTACTTC GTCAGGGCTC 900
ACGCTCTACT AAGGATGTGC ACCATGGTAA GGCATCTCGC GGGGGGTAGG TACGTCCAGA 960
TGGTGCTACT AGCCCTTGGC AGGTGGACTG GCACCTACAT CTATGACCAC CTCACCCCTA 1020
TGTGGGATTG GGCTGCTAAT GGCCTGCGGG ACTTGGCGGT GCGCGTGGAG CCTATCATCT 1080
TCAGTCCGAT GGAGAAAAA GTCATCGTCT GGGGAGCGGA GACAGCTGCT TGCGGGGATA 1140
TCTTACACGG ACTTCCCGTG TCCGCCCGAC TTGGCCGGGA GGTCTCTCTT GGCCCAAGCTG 1200

```

ATGGCTATAC CTCCAAGGG TGGAGTCTTC TGCCCCCAT CACTGCTTAT GCCCAGCAGA 1260
 CACGCCGCCCT TTTGGGCACC ATAGTGGTGA GCATGACGGG GCGCGACAAG ACAGAACAGG 1320
 CCGGGGAGAT TCAGGTCTG TCCACGGTCA CTCAGTCTTT CCTCGGAACA ACCATCTCGG 1380
 GGGTCTTATG GACTGTCTAC CATGGAGCTG GCAACAAGAC TCTAGCCGGC TCACGGGGTC 1440
 CGGTACACA GATGTACTCC AGTGTGAGG GGGACTTAGT GGGGTGGCCC AGCCCCCGG 1500
 GGACCAATC TTTGGAGCCG TGCAGTGTG GAGCGGTGGA CCTATACCTG GTCACGCGAA 1560
 ACGTGATGT CATCCCGGT CGAAGACCG GGCACAAGCG AGGAGCGCTA CTCTCCCCGA 1620
 GACCTCTTTC CACCTTGAAG GGTCTCTCGG GGGGCCCGGT GCTCTGCCCC AGAGGCCACG 1680
 CTGTGCGGGT CTTCGGGCA GCCGTGTCT CCCGGGGCGT GGGCAAGTCC ATAGATTTTA 1740
 TCCCCGTTGA GACACTTGAC ATCGTCACTC GGTCCCCAC CTTTAGTGAC AACAGCACAC 1800
 CACCTGCTGT CCCCCAACT TATCAGGTG GGTACTTACA TGCCCCGACT GGTAGTGGAA 1860
 AGAGCACCA AGTCCTGTG GGTATGCGG CTCAGGGGTA CAAAGTGCTA GTGCTTAATC 1920
 CCTGGGTGGC TGCCACCCTG GGGTTTGGG CGTACTGTG CAAGGCACAT GGCATCAATC 1980
 CCAACATTAG GACTGGGGT AGGACTGTGA CGACCGGGG GCCCATCAG TACTCCACAT 2040
 ATGGCAAAT CCTCGCGAT GGGGGTGGC CAGGCGGCG CTATGACATC ATCATATGCG 2100
 ATGAATGCCA TGCCGTGGAC TCTACCACCA TTCTCGCAT CGGAACAGTC CTCGATCAAG 2160
 CAGAGACAGC CGGGGTCAG CTAAGTGTG TGGTACGGC TACGCCCCC GGGTCAGTGA 2220
 CAACCCCCA CCCCCACATA GAGGAGGTG CCCTCGGGCA GGAGGGTGAG ATCCCCTTCT 2280
 ATGGGAGGGC GATTCCCCTG TCATACATCA AGGGAGGAAG ACACTTGATC TTCTGCCACT 2340
 CAAAGAAAAA GTGTGACGAG CTGCGGGCG CCCTTCGGG TATGGGCTG AACGCAGTGG 2400
 CATACTACAG AGGGCTGGAC GTCTCCGTAA TACCAACTCA GGGAGACGTA GTGGTCGTG 2460
 CCACCGACGC CCTCATGACG GGGTTTACTG GAGACTTTGA CTCGTGATC GACTGCAACG 2520
 TAGCGGTCA TCAAGTTGTA GACTTCAGCT TGGACCCAC ATTCACCATA ACCACACAGA 2580
 CTGTCCCTCA AGAGGTGTG TCAGTAGGC AGCGCCGGG CCGCACGGG AGGGGAGAC 2640
 TGGGTATTA TAGGTATGTT TCCACTGGT AGCGAGCCTC AGGAATGTT GACAGTGTAG 2700
 TGCTCTGCGA GTGCTACGAT GCAGGGGCG CATGGTATGA GCTCACACCA GCGGAGACCA 2760
 CCGTCAGGCT CAGAGCATAT TTCAACACAC CTGGTTTGG TGTGTGCCAA GACCATCTTG 2820
 AGTTTGGGA GCAGTTTC ACCGGCCTCA CACACATAGA TGCCCACTTC CTTTCCCAA 2880
 CAAAGCAATC GGGGAAAAAT TTCGCATACT TAACAGCCTA CCAGGCTACA GTGTGGCTA 2940

GGGCCAAAGC CCCCCCCCCG TCCTGGGACG TCATGTGGAA GTGTTTGA CTCTCAAGC 3000
CCACACTCGT GGGCCCCACA CCTCTCTGT ACCGCTTGGG CTCTGTTACC AACGAGGTCA 3060
CCCTCAGGCA TCCTGTGACG AAATACATCG CCACCTGCAT GCAAGCCGAC CTGAGGTCA 3120
TGACCAGCAC GTGGGTCTTA GCTGGGGGGG TCTTGGCGG CGTCGCCGG TACTGCCTGG 3180
CGACCGGGTG TGTTCGATC ATCGGCGGT TGCAGTTAA CCAGCGAGCC GTGTTGAC 3240
CGGACAAGGA GGTCTCTAT GAGGCITTTG ATGAGATGGA GGAATGTGC TCTAGAGCGG 3300
CTCTCAITGA AGAGGGGACG CGGATAGCCG AGATGCTGAA GTCCAAGATC CAAGGCTTAT 3360
TGCAGCAAGC TTCCAAACAA GCTCAAGACA TACAACCCGC TGTGCAGGCT TCTTGGCCCA 3420
AGGTAGAGCA ATTCTGGGC AAACACATGT GGAACCTCAT CAGCGGCATT CAATACCTCG 3480
CAGGACTATC AACACTGCCA GGAACCCCTG CTGTAGCTTC CATGATGGCA TTCAGTGGCG 3540
CCCTCACCAG TCGTTGTCA ACTAGCACC CTATCCTTCT CAACATTTG GGGGGCTGGC 3600
TAGCATCCCA AATTGCGCT CCCGCGGGG CTACCGGCTT CGTCGTAGT GGCCTGGTGG 3660
GGGTGCGT AGGCAGCATA GGTTCGGTA AGGTGCTGGT GGACATCTG GCAGGGTATG 3720
GTGCGGGCAT TTGCGGGCT CTGTCGCAT TCAAGATCAT GTCTGGCGAG AAGCCCTCCA 3780
TGGAGGATGT TGTCAACCTG CTGCCTGGAA TTCTGTCTCC GGGTGGCCTG GTGGTGGGAG 3840
TCATCTGCGC GGCCATCTG CCGGACACG TGGGACCGG GGAAGGCGCT GTCCAATGGA 3900
TGAATAGGCT CATTGCCCTT GCTTCCAGAG GAAACACGT CGCCCCACC CACTACGTGA 3960
CGGAGTCGGA 3970

配列番号 : 4

配列の長さ : 2693

配列の型 : 核酸

鎖の数 : 一本鎖

トポロジー : 直鎖状

配列の種類 : cDNA to genomic RNA

特徴を決定した方法 : E

配列

```

ATTCTGTCTC CGGGTGCCCT GGTGGTGGGA GTCATCTGGC CGGCCATCCT GCGCCGACAC   60
GTGGGACCGG GGAAGGCGC TGTCCAATGG ATGAATAGGC TCATTGCCIT TGCTTCAGA 120
GGAAACCACG TCGCCCCAC CCACTACGTG ACGGAGTCGG ATGCGTCGCA GCGTGTGACC 180
CAACTACTTG GCTCCCTTAC CATAACCAGC CTGCTCAGGA GACTCCACAA CTGGATTACT 240
GAAGACTGCC CCATCCCATG CAGCGGCTCG TGGCTCCGCG ATGTGTGGGA TTGGGTTTGC 300
ACCATCCTAA CAGACTTTAA AAAGTGGCTG ACCTCCAAAT TGTCCCAAAT GATGCCCTGGT 360
CTCCCCTTTA TCTCTTGTC AAAGGGGTAC AAGGCGGTGT GGGCTGGCAC TGGTATCATG 420
ACCACAGGGT GTCCTTGCGG CGCCAATATC TCTGGCAATG TCCGCTGGG CTCCATGAGA 480
ATTACGGGGC CCAAAAGCTG CATGAATATC TGGCAGGGGA CCTTCCCAT CAATTGTTAC 540
ACGGAGGGCC AGTGGGTGCC GAAACCCGCA CCAAACTTTA AGATCGCCAT CTGGAGGGTG 600
GCGGCCTCAG AGTACGCGGA GGTGACGCAG CACGGGTCAT ACCACTACAT AACAGGACTT 660
ACCACTGATA ACTTGAAAGT TCCTTGCCAA CTACCTTCTC CAGAGTTCTT TTCCTGGGTG 720
GACGGAGTGC AGATCCATAG GTTTGCCCCC ATACCGAAGC CGTTTTTTCG GGATGAGGTC 780
TCGTTCTGCG TTGGGCTTAA TTCATTTGTC GTCGGGCTC AGCTCCCTTG CGATCCTGAA 840
CCTGACACAG ACGTATTGAC GTCCATGCTA ACAGACCCAT CCCATATCAC GCGGAGACT 900
GCAGCGCGGC GTTTGGCAGG GGGGTACCC CCGTCCGAGG CAAGCTCCTC AGCGAGCCAG 960
CTATCGGCAC CATCGCTGCG AGCCACCTGC ACCACCCAGG GCAAGGCCTA TGATGTGGAC 1020
ATGTTGGATG CCAACCTGTT CATGGGGGCG GATGTGACCC GGATAGAGTC TGAGTCCAAA 1080
GTGGTCGTTT TGGACTCTCT CGACCCAATG GTCGAAGAAA GGAGCGACCT TGAGCCTTCG 1140
ATACCATCGG AATATATGCT CCCCAAGAAG AGATTCCAC CAGCCTTACC GGCTTGGGCA 1200

```

CGGCCTGATT ACAACCCACC GCTTGTGGAA TCGTGGAGA GGCCAGATTA CCAACCGGCC 1260
 ACTGTTGCGG GCTGCGCTCT CCCCCCCTT AAGAAAACCC CGACGCCTCC CCCAAGGAGA 1320
 CGCCGGACAG TGGGTCTGAG TGAGAGCTCC ATAGCAGATG CCTACAACA GCTGGCCATC 1380
 AAGTCCTTTG GCCAGCCCCC CCCAAGCGGC GATTGAGGCC TTTCACGGG GGCAGGACGA 1440
 GCCGATTCCG GCAGTCGGAC GCGCCCGCAT GAGTTGGGCC TTTCGGAGAC AGGTTCATC 1500
 TCCTCCATGC CCCCTCTCGA GGGGGAGCCT GGAGATCCAG ACTTGGAGCC TGAGCAGGTA 1560
 GAGCTTCAAC CTCCCCCCA GGGGGGGGTG GTAACCCCGG GCTCAGGCTC GGGGTCTTGG 1620
 TCTACTTGCT CCGAGGAGGA CGACTCCGTC GTGTGCTGCT CCATGTCATA CTCCTGGACC 1680
 GGGGTCTTAA TAACTCCTTG TAGCCCCGAA GAGGAAAAGT TGCCAATTGG CCCCTTGAGC 1740
 AACTCCCTGT TGCGATATCA CAACAAGGTG TACTGTACCA CATCAAAGAG CGCCTCATT 1800
 AGGGCTAAAA AGGTAACITT TGATAGGATG CAAGCGCTCG ACGCTCATT 1860
 TTGAAGGACA TTAAGCTAGC GGCCTCCAAG GTCACCGCAA GGCTTCTCAC TTAGAGGAG 1920
 GCCTGCCAGT TAACTCCACC CCCTCTGCA AGATCCAAGT ATGGGTTTGG GGCTAAGGAG 1980
 GTCCGCAGCT TGTCGGGAG AGCCGTTAAC CACATCAAGT CCGTGTGGAA GGACCTCTG 2040
 GAAGACACAC AAACACCAAT TCCTACAACC ATCATGGCCA AAAATCAGGT GTTCTGCGTG 2100
 GACCCACCA AGGGGGGTAA GAAAGCAGCT CGCCTTATCG TTTACCTGA CCTCGGCGTC 2160
 AGGGTCTGCG AGAAAATGGC CCTTTATGAT ATCAGACAAA AGCTTCCTCA GCGGTGATG 2220
 GGGGCTTCTT ATGGATTCCA GTACTCCCC GCTCAGCGGG TGGAGTTCT CTTGAAGGCA 2280
 TGGGCGGAAA AGAAAGACCC TATGGGTTT TCGTATGATA CCGATGCTT TGAATCAACC 2340
 CTCCTGAGA GAGACATCAG GACTGAGGAG TCCATATATC GGGCTTGTTC CTTGCCCGAG 2400
 GAGGCCACAC CTGCCATACA CTCCTGACT GAGAGACTTT ACCTGGGAGG GCCCATGTT 2460
 AACAGCAAGG GCCAGACTG CCGGTACAGG CGTTGCCCGG CCAGCGGGGT GCTTACCACT 2520
 AGCATGGGGA ACACCATCAC ATGCTATGTG AAAGCCTTAG CGCCTGTAA GCTGCGAGG 2580
 ATAATTGCGC CCACAATGCT GGTATCGGC GATGACTTGG TTGTCATCTC AGAGAGCCAG 2640
 GGGACCGAGG AGGACGAGCG GAACCTGAGA GCCTTCACGG AGGCTATGAC CAG 2693

配列番号 : 5

配列の長さ : 3033

配列の型 : アミノ酸

トポロジー : 直鎖状

配列の種類 : 蛋白質

配列

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr		
	5	10
Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly Gln Ile		15
	20	25
Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg Leu Gly		30
	35	40
Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser Gln Pro Arg Gly		45
	50	55
Arg Arg Gln Pro Ile Pro Lys Asp Arg Arg Ser Thr Gly Lys Ser		60
	65	70
Trp Gly Lys Pro Gly Tyr Pro Trp Pro Leu Tyr Gly Asn Glu Gly		75
	80	85
Leu Gly Trp Ala Gly Trp Leu Leu Ser Pro Arg Gly Ser Arg Pro		90
	95	100
Ser Trp Gly Pro Asn Asp Pro Arg His Arg Ser Arg Asn Val Gly		105
	110	115
Lys Val Ile Asp Thr Leu Thr Cys Gly Phe Ala Asp Leu Met Gly		120
	125	130
Tyr Ile Pro Val Val Gly Ala Pro Leu Gly Gly Val Ala Arg Ala		135
	140	145
Leu Ala His Gly Val Arg Val Leu Glu Asp Gly Val Asn Phe Ala		150
	155	160
		165

Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala		
	170	175 180
Leu Leu Ser Cys Ile Thr Thr Pro Val Ser Ala Ala Glu Val Lys		
	185	190 195
Asn Ile Ser Thr Gly Tyr Met Val Thr Asn Asp Cys Thr Asn Asp		
	200	205 210
Ser Ile Thr Trp Gln Leu Gln Ala Ala Val Leu His Val Pro Gly		
	215	220 225
Cys Val Pro Cys Glu Lys Val Gly Asn Thr Ser Arg Cys Trp Ile		
	230	235 240
Pro Val Ser Pro Asn Val Ala Val Gln Gln Pro Gly Ala Leu Thr		
	245	250 255
Gln Gly Leu Arg Thr His Ile Asp Met Val Val Met Ser Ala Thr		
	260	265 270
Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys Gly Gly Val Met		
	275	280 285
Leu Ala Ala Gln Met Phe Ile Val Ser Pro Gln His His Trp Phe		
	290	295 300
Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly Thr Ile Thr Gly		
	305	310 315
His Arg Met Ala Trp Asp Met Met Met Asn Trp Ser Pro Thr Ala		
	320	325 330
Thr Met Ile Leu Ala Tyr Ala Met Arg Val Pro Glu Val Ile Ile		
	335	340 345
Asp Ile Ile Gly Gly Ala His Trp Gly Val Met Phe Gly Leu Ala		
	350	355 360
Tyr Phe Ser Met Gln Gly Ala Trp Ala Lys Val Val Val Ile Leu		
	365	370 375
Leu Leu Ala Ala Gly Val Asp Ala Gln Thr His Thr Val Gly Gly		

380	385	390
Ser Thr Ala His Asn Ala Arg Thr Leu Thr Gly Met Phe Ser Leu		
395	400	405
Gly Ala Arg Gln Lys Ile Gln Leu Ile Asn Thr Asn Gly Ser Trp		
410	415	420
His Ile Asn Arg Thr Ala Leu Asn Cys Asn Asp Ser Leu His Thr		
425	430	435
Gly Phe Leu Ala Ser Leu Phe Tyr Thr His Ser Phe Asn Ser Ser		
440	445	450
Gly Cys Pro Glu Arg Met Ser Ala Cys Arg Ser Ile Glu Ala Phe		
455	460	465
Arg Val Gly Trp Gly Ala Leu Gln Tyr Glu Asp Asn Val Thr Asn		
470	475	480
Pro Glu Asp Met Arg Pro Tyr Cys Trp His Tyr Pro Pro Arg Gln		
485	490	495
Cys Gly Val Val Ser Ala Ser Ser Val Cys Gly Pro Val Tyr Cys		
500	505	510
Phe Thr Pro Ser Pro Val Val Val Gly Thr Thr Asp Arg Leu Gly		
515	520	525
Ala Pro Thr Tyr Thr Trp Gly Glu Asn Glu Thr Asp Val Phe Leu		
530	535	540
Leu Asn Ser Thr Arg Pro Pro Gln Gly Ser Trp Phe Gly Cys Thr		
545	550	555
Trp Met Asn Ser Thr Gly Tyr Thr Lys Thr Cys Gly Ala Pro Pro		
560	565	570
Cys Arg Ile Arg Ala Asp Phe Asn Ala Ser Met Asp Leu Leu Cys		
575	580	585
Pro Thr Asp Cys Phe Arg Lys His Pro Asp Thr Thr Tyr Ile Lys		
590	595	600

Cys Gly Ser Gly Pro Trp Leu Thr Pro Arg Cys Leu Ile Asp Tyr		
605	610	615
Pro Tyr Arg Leu Trp His Tyr Pro Cys Thr Val Asn Tyr Thr Ile		
620	625	630
Phe Lys Ile Arg Met Tyr Val Gly Gly Val Glu His Arg Leu Thr		
635	640	645
Ala Ala Cys Asn Phe Thr Arg Gly Asp Arg Cys Asn Leu Glu Asp		
650	655	660
Arg Asp Arg Ser Gln Leu Ser Pro Leu Leu His Ser Thr Thr Glu		
665	670	675
Trp Ala Ile Leu Pro Cys Thr Tyr Ser Asp Leu Pro Ala Leu Ser		
680	685	690
Thr Gly Leu Leu His Leu His Gln Asn Ile Val Asp Val Gln Phe		
695	700	705
Met Tyr Gly Leu Ser Pro Ala Leu Thr Lys Tyr Ile Val Arg Trp		
710	715	720
Glu Trp Val Val Leu Leu Phe Leu Leu Leu Ala Asp Ala Arg Val		
725	730	735
Cys Ala Cys Leu Trp Met Leu Ile Leu Leu Gly Gln Ala Glu Ala		
740	745	750
Ala Leu Glu Lys Leu Val Val Leu His Ala Ala Ser Ala Ala Ser		
755	760	765
Cys Asn Gly Phe Leu Tyr Phe Val Ile Phe Phe Val Ala Ala Trp		
770	775	780
Tyr Ile Lys Gly Arg Val Val Pro Leu Ala Thr Tyr Ser Leu Thr		
785	790	795
Gly Leu Trp Ser Phe Gly Leu Leu Leu Leu Ala Leu Pro Gln Gln		
800	805	810
Ala Tyr Ala Tyr Asp Ala Ser Val His Gly Gln Ile Gly Ala Ala		

815	820	825
Leu Leu Val Leu Ile Thr Leu Phe Thr	Leu Thr Pro Gly Tyr Lys	
830	835	840
Thr Leu Leu Ser Arg Phe Leu Trp Trp	Leu Cys Tyr Leu Leu Thr	
845	850	855
Leu Ala Glu Ala Met Val Gln Glu Trp	Ala Pro Pro Met Gln Val	
860	865	870
Arg Gly Gly Arg Asp Gly Ile Ile Trp	Ala Val Ala Ile Phe Cys	
875	880	885
Pro Gly Val Val Phe Asp Ile Thr Lys	Trp Leu Leu Ala Val Leu	
890	895	900
Gly Pro Ala Tyr Leu Leu Lys Gly Ala	Leu Thr Arg Val Pro Tyr	
905	910	915
Phe Val Arg Ala His Ala Leu Leu Arg	Met Cys Thr Met Val Arg	
920	925	930
His Leu Ala Gly Gly Arg Tyr Val Gln	Met Val Leu Leu Ala Leu	
935	940	945
Gly Arg Trp Thr Gly Thr Tyr Ile Tyr	Asp His Leu Thr Pro Met	
950	955	960
Ser Asp Trp Ala Ala Asn Gly Leu Arg	Asp Leu Ala Val Ala Val	
965	970	975
Glu Pro Ile Ile Phe Ser Pro Met Glu	Lys Lys Val Ile Val Trp	
980	985	990
Gly Ala Glu Thr Ala Ala Cys Gly Asp	Ile Leu His Gly Leu Pro	
995	1000	1005
Val Ser Ala Arg Leu Gly Arg Glu Val	Leu Leu Gly Pro Ala Asp	
1010	1015	1020
Gly Tyr Thr Ser Lys Gly Trp Ser Leu	Leu Ala Pro Ile Thr Ala	
1025	1030	1035

Tyr Ala Gln Gln Thr Arg Gly Leu Leu Gly Thr Ile Val Val Ser		
1040	1045	1050
Met Thr Gly Arg Asp Lys Thr Glu Gln Ala Gly Glu Ile Glu Val		
1055	1060	1065
Leu Ser Thr Val Thr Gln Ser Phe Leu Gly Thr Thr Ile Ser Gly		
1070	1075	1080
Val Leu Trp Thr Val Tyr His Gly Ala Gly Asn Lys Thr Leu Ala		
1085	1090	1095
Gly Ser Arg Gly Pro Val Thr Gln Met Tyr Ser Ser Ala Glu Gly		
1100	1105	1110
Asp Leu Val Gly Trp Pro Ser Pro Pro Gly Thr Lys Ser Leu Glu		
1115	1120	1125
Pro Cys Thr Cys Gly Ala Val Asp Leu Tyr Leu Val Thr Arg Asn		
1130	1135	1140
Ala Asp Val Ile Pro Ala Arg Arg Arg Gly Asp Lys Arg Gly Ala		
1145	1150	1155
Leu Leu Ser Pro Arg Pro Leu Ser Thr Leu Lys Gly Ser Ser Gly		
1160	1165	1170
Gly Pro Val Leu Cys Pro Arg Gly His Ala Val Gly Val Phe Arg		
1175	1180	1185
Ala Ala Val Cys Ser Arg Gly Val Ala Lys Ser Ile Asp Phe Ile		
1190	1195	1200
Pro Val Glu Thr Leu Asp Ile Val Thr Arg Ser Pro Thr Phe Ser		
1205	1210	1215
Asp Asn Ser Thr Pro Pro Ala Val Pro Gln Thr Tyr Gln Val Gln		
1220	1225	1230
Tyr Leu His Ala Pro Thr Gly Ser Gly Lys Ser Thr Lys Val Pro		
1235	1240	1245
Val Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro		

1250	1255	1260
Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Leu Ser Lys Ala		
1265	1270	1275
His Gly Ile Asn Pro Asn Ile Arg Thr Gly Val Arg Thr Val Thr		
1280	1285	1290
Thr Gly Ala Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala		
1295	1300	1305
Asp Gly Gly Cys Ala Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp		
1310	1315	1320
Glu Cys His Ala Val Asp Ser Thr Thr Ile Leu Gly Ile Gly Thr		
1325	1330	1335
Val Leu Asp Gln Ala Glu Thr Ala Gly Val Arg Leu Thr Val Leu		
1340	1345	1350
Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Thr Pro His Pro Asn		
1355	1360	1365
Ile Glu Glu Val Ala Leu Gly Gln Glu Gly Glu Ile Pro Phe Tyr		
1370	1375	1380
Gly Arg Ala Ile Pro Leu Ser Tyr Ile Lys Gly Gly Arg His Leu		
1385	1390	1395
Ile Phe Cys His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Ala		
1400	1405	1410
Leu Arg Gly Met Gly Leu Asn Ala Val Ala Tyr Tyr Arg Gly Leu		
1415	1420	1425
Asp Val Ser Val Ile Pro Thr Gln Gly Asp Val Val Val Val Ala		
1430	1435	1440
Thr Asp Ala Leu Met Thr Gly Phe Thr Gly Asp Phe Asp Ser Val		
1445	1450	1455
Ile Asp Cys Asn Val Ala Val Thr Gln Val Val Asp Phe Ser Leu		
1460	1465	1470

Asp Pro Thr Phe Thr Ile Thr Thr Gln Thr Val Pro Gln Asp Ala	1475	1480	1485
Val Ser Arg Ser Gln Arg Arg Gly Arg Thr Gly Arg Gly Arg Leu	1490	1495	1500
Gly Ile Tyr Arg Tyr Val Ser Thr Gly Glu Arg Ala Ser Gly Met	1505	1510	1515
Phe Asp Ser Val Val Leu Cys Glu Cys Tyr Asp Ala Gly Ala Ala	1520	1525	1530
Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg Ala	1535	1540	1545
Tyr Phe Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu	1550	1555	1560
Phe Trp Glu Ala Val Phe Thr Gly Leu Thr His Ile Asp Ala His	1565	1570	1575
Phe Leu Ser Gln Thr Lys Gln Ser Gly Glu Asn Phe Ala Tyr Leu	1580	1585	1590
Thr Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Lys Ala Pro Pro	1595	1600	1605
Pro Ser Trp Asp Val Met Trp Lys Cys Leu Thr Arg Leu Lys Pro	1610	1615	1620
Trp Leu Val Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ser Val	1625	1630	1635
Thr Asn Glu Val Thr Leu Thr His Pro Val Thr Lys Tyr Ile Ala	1640	1645	1650
Thr Cys Met Gln Ala Asp Leu Glu Val Met Thr Ser Thr Trp Val	1655	1660	1665
Leu Ala Gly Gly Val Leu Ala Ala Val Ala Ala Tyr Cys Leu Ala	1670	1675	1680
Thr Gly Cys Val Cys Ile Ile Gly Arg Leu His Val Asn Gln Arg			

1685	1690	1695
Ala Val Val Ala Pro Asp Lys Glu Val Leu Tyr Glu Ala Phe Asp		
1700	1705	1710
Glu Met Glu Glu Cys Ala Ser Arg Ala Ala Leu Ile Glu Glu Gly		
1715	1720	1725
Gln Arg Ile Ala Glu Met Leu Lys Ser Lys Ile Gln Gly Leu Leu		
1730	1735	1740
Gln Gln Ala Ser Lys Gln Ala Gln Asp Ile Gln Pro Ala Val Gln		
1745	1750	1755
Ala Ser Trp Pro Lys Val Glu Gln Phe Trp Ala Lys His Met Trp		
1760	1765	1770
Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu		
1775	1780	1785
Pro Gly Asn Pro Ala Val Ala Ser Met Met Ala Phe Ser Ala Ala		
1790	1795	1800
Leu Thr Ser Pro Leu Ser Thr Ser Thr Thr Ile Leu Leu Asn Ile		
1805	1810	1815
Leu Gly Gly Trp Leu Ala Ser Gln Ile Ala Pro Pro Ala Gly Ala		
1820	1825	1830
Thr Gly Phe Val Val Ser Gly Leu Val Gly Ala Ala Val Gly Ser		
1835	1840	1845
Ile Gly Leu Gly Lys Val Leu Val Asp Ile Leu Ala Gly Tyr Gly		
1850	1855	1860
Ala Gly Ile Ser Gly Ala Leu Val Ala Phe Lys Ile Met Ser Gly		
1865	1870	1875
Glu Lys Pro Ser Met Glu Asp Val Val Asn Leu Leu Pro Gly Ile		
1880	1885	1890
Leu Ser Pro Gly Ala Leu Val Val Gly Val Ile Cys Ala Ala Ile		
1895	1900	1905

Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala Val Gln Trp Met		
1910	1915	1920
Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His Val Ala Pro		
1925	1930	1935
Thr His Tyr Val Thr Glu Ser Asp Ala Ser Gln Arg Val Thr Gln		
1940	1945	1950
Leu Leu Gly Ser Leu Thr Ile Thr Ser Leu Leu Arg Arg Leu His		
1955	1960	1965
Asn Trp Ile Thr Glu Asp Cys Pro Ile Pro Cys Ser Gly Ser Trp		
1970	1975	1980
Leu Arg Asp Val Trp Asp Trp Val Cys Thr Ile Leu Thr Asp Phe		
1985	1990	1995
Lys Asn Trp Leu Thr Ser Lys Leu Phe Pro Lys Met Pro Gly Leu		
2000	2005	2010
Pro Phe Ile Ser Cys Gln Lys Gly Tyr Lys Gly Val Trp Ala Gly		
2015	2020	2025
Thr Gly Ile Met Thr Thr Arg Cys Pro Cys Gly Ala Asn Ile Ser		
2030	2035	2040
Gly Asn Val Arg Leu Gly Ser Met Arg Ile Thr Gly Pro Lys Thr		
2045	2050	2055
Cys Met Asn Ile Trp Gln Gly Thr Phe Pro Ile Asn Cys Tyr Thr		
2060	2065	2070
Glu Gly Gln Cys Val Pro Lys Pro Ala Pro Asn Phe Lys Ile Ala		
2075	2080	2085
Ile Trp Arg Val Ala Ala Ser Glu Tyr Ala Glu Val Thr Gln His		
2090	2095	2100
Gly Ser Tyr His Tyr Ile Thr Gly Leu Thr Thr Asp Asn Leu Lys		
2105	2110	2115
Val Pro Cys Gln Leu Pro Ser Pro Glu Phe Phe Ser Trp Val Asp		

2120	2125	2130
Gly Val Gln Ile His Arg Phe Ala Pro Ile Pro Lys Pro Phe Phe		
2135	2140	2145
Arg Asp Glu Val Ser Phe Cys Val Gly Leu Asn Ser Phe Val Val		
2150	2155	2160
Gly Ser Gln Leu Pro Cys Asp Pro Glu Pro Asp Thr Asp Val Leu		
2165	2170	2175
Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu Thr Ala		
2180	2185	2190
Ala Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Glu Ala Ser Ser		
2195	2200	2205
Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Arg Ala Thr Cys Thr		
2210	2215	2220
Thr His Gly Lys Ala Tyr Asp Val Asp Met Val Asp Ala Asn Leu		
2225	2230	2235
Phe Met Gly Gly Asp Val Thr Arg Ile Glu Ser Glu Ser Lys Val		
2240	2245	2250
Val Val Leu Asp Ser Leu Asp Pro Met Val Glu Glu Arg Ser Asp		
2255	2260	2265
Leu Glu Pro Ser Ile Pro Ser Glu Tyr Met Leu Pro Lys Lys Arg		
2270	2275	2280
Phe Pro Pro Ala Leu Pro Ala Trp Ala Arg Pro Asp Tyr Asn Pro		
2285	2290	2295
Pro Leu Val Glu Ser Trp Lys Arg Pro Asp Tyr Gln Pro Ala Thr		
2300	2305	2310
Val Ala Gly Cys Ala Leu Pro Pro Pro Lys Lys Thr Pro Thr Pro		
2315	2320	2325
Pro Pro Arg Arg Arg Arg Thr Val Gly Leu Ser Glu Ser Ser Ile		
2330	2335	2340

Ala Asp Ala Leu Gln Gln Leu Ala Ile Lys Ser Phe Gly Gln Pro			
2345	2350	2355	
Pro Pro Ser Gly Asp Ser Gly Leu Ser Thr Gly Ala Asp Ala Ala			
2360	2365	2370	
Asp Ser Gly Ser Arg Thr Pro Pro Asp Glu Leu Ala Leu Ser Glu			
2375	2380	2385	
Thr Gly Ser Ile Ser Ser Met Pro Pro Leu Glu Gly Glu Pro Gly			
2390	2395	2400	
Asp Pro Asp Leu Glu Pro Glu Gln Val Glu Leu Gln Pro Pro Pro			
2405	2410	2415	
Gln Gly Gly Val Val Thr Pro Gly Ser Gly Ser Gly Ser Trp Ser			
2420	2425	2430	
Thr Cys Ser Glu Glu Asp Asp Ser Val Val Cys Cys Ser Met Ser			
2435	2440	2445	
Tyr Ser Trp Thr Gly Ala Leu Ile Thr Pro Cys Ser Pro Glu Glu			
2450	2455	2460	
Glu Lys Leu Pro Ile Asn Pro Leu Ser Asn Ser Leu Leu Arg Tyr			
2465	2470	2475	
His Asn Lys Val Tyr Cys Thr Thr Ser Lys Ser Ala Ser Leu Arg			
2480	2485	2490	
Ala Lys Lys Val Thr Phe Asp Arg Met Gln Ala Leu Asp Ala His			
2495	2500	2505	
Tyr Asp Ser Val Leu Lys Asp Ile Lys Leu Ala Ala Ser Lys Val			
2510	2515	2520	
Thr Ala Arg Leu Leu Thr Leu Glu Glu Ala Cys Gln Leu Thr Pro			
2525	2530	2535	
Pro His Ser Ala Arg Ser Lys Tyr Gly Phe Gly Ala Lys Glu Val			
2540	2545	2550	

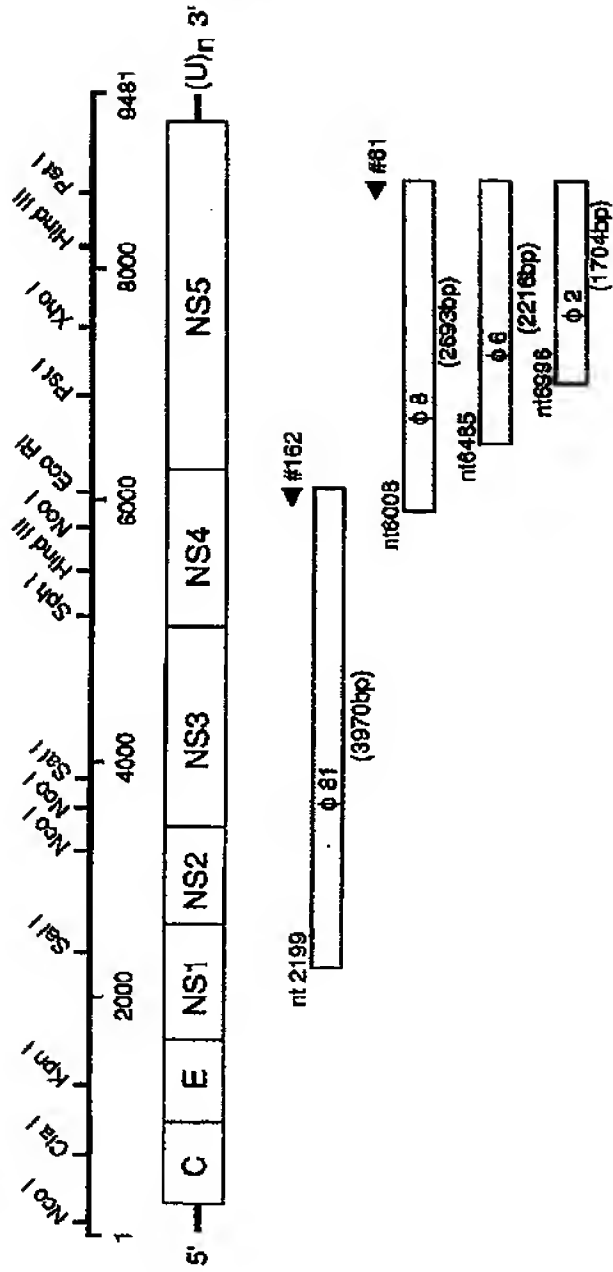
Arg Ser Leu Ser Gly Arg Ala Val Asn His Ile Lys Ser Val Trp			
2555	2560	2565	
Lys Asp Leu Leu Glu Asp Thr Gln Thr Pro Ile Pro Thr Thr Ile			
2570	2575	2580	
Met Ala Lys Asn Glu Val Phe Cys Val Asp Pro Thr Lys Gly Gly			
2585	2590	2595	
Lys Lys Ala Ala Arg Leu Ile Val Tyr Pro Asp Leu Gly Val Arg			
2600	2605	2610	
Val Cys Glu Lys Met Ala Leu Tyr Asp Ile Thr Gln Lys Leu Pro			
2615	2620	2625	
Gln Ala Val Met Gly Ala Ser Tyr Gly Phe Gln Tyr Ser Pro Ala			
2630	2635	2640	
Gln Arg Val Glu Phe Leu Leu Lys Ala Trp Ala Glu Lys Lys Asp			
2645	2650	2655	
Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val			
2660	2665	2670	
Thr Glu Arg Asp Ile Arg Thr Glu Glu Ser Ile Tyr Arg Ala Cys			
2675	2680	2685	
Ser Leu Pro Glu Glu Ala His Thr Ala Ile His Ser Leu Thr Glu			
2690	2695	2700	
Arg Leu Tyr Val Gly Gly Pro Met Phe Asn Ser Lys Gly Gln Thr			
2705	2710	2715	
Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser			
2720	2725	2730	
Met Gly Asn Thr Ile Thr Cys Tyr Val Lys Ala Leu Ala Ala Cys			
2735	2740	2745	
Lys Ala Ala Gly Ile Ile Ala Pro Thr Met Leu Val Cys Gly Asp			
2750	2755	2760	

Asp Leu Val Val Ile Ser Glu Ser Gln Gly Thr Glu Glu Asp Glu			
2765	2770	2775	
Arg Asn Leu Arg Ala Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala			
2780	2785	2790	
Pro Pro Gly Asp Pro Pro Arg Pro Glu Tyr Asp Leu Glu Leu Ile			
2795	2800	2805	
Thr Ser Cys Ser Ser Asn Val Ser Val Ala Leu Gly Pro Gln Gly			
2810	2815	2820	
Arg Arg Arg Tyr Tyr Leu Thr Arg Asp Pro Thr Thr Pro Ile Ala			
2825	2830	2835	
Arg Ala Ala Trp Glu Thr Val Arg His Ser Pro Val Asn Ser Trp			
2840	2845	2850	
Leu Gly Asn Ile Ile Gln Tyr Ala Pro Thr Ile Trp Ala Arg Met			
2855	2860	2865	
Val Leu Met Thr His Phe Phe Ser Ile Leu Met Ala Gln Asp Thr			
2870	2875	2880	
Leu Asp Gln Asn Leu Asn Phe Glu Met Tyr Gly Ala Val Tyr Ser			
2885	2890	2895	
Val Ser Pro Leu Asp Leu Pro Ala Ile Ile Glu Arg Leu His Gly			
2900	2905	2910	
Leu Asp Ala Phe Ser Leu His Thr Tyr Thr Pro His Glu Leu Thr			
2915	2920	2925	
Arg Val Ala Ser Ala Leu Arg Lys Leu Gly Ala Pro Pro Leu Arg			
2930	2935	2940	
Ala Trp Lys Ser Arg Ala Arg Ala Val Arg Ala Ser Leu Ile Ser			
2945	2950	2955	
Arg Gly Gly Arg Ala Ala Val Cys Gly Arg Tyr Leu Phe Asn Trp			
2960	2965	2970	

Ala Val Lys Thr Lys Leu Lys Leu Thr Pro Leu Pro Glu Ala Arg			
	2975	2980	2985
Leu Leu Asp Leu Ser Ser Trp Phe Thr Val Gly Ala Gly Gly Gly			
	2990	2995	3000
Asp Ile Tyr His Ser Val Ser Arg Ala Arg Pro Arg Leu Leu Leu			
	3005	3010	3015
Leu Gly Leu Leu Leu Leu Phe Val Gly Val Gly Leu Phe Leu Leu			
	3020	3025	3030
Pro Ala Arg			
	3033		

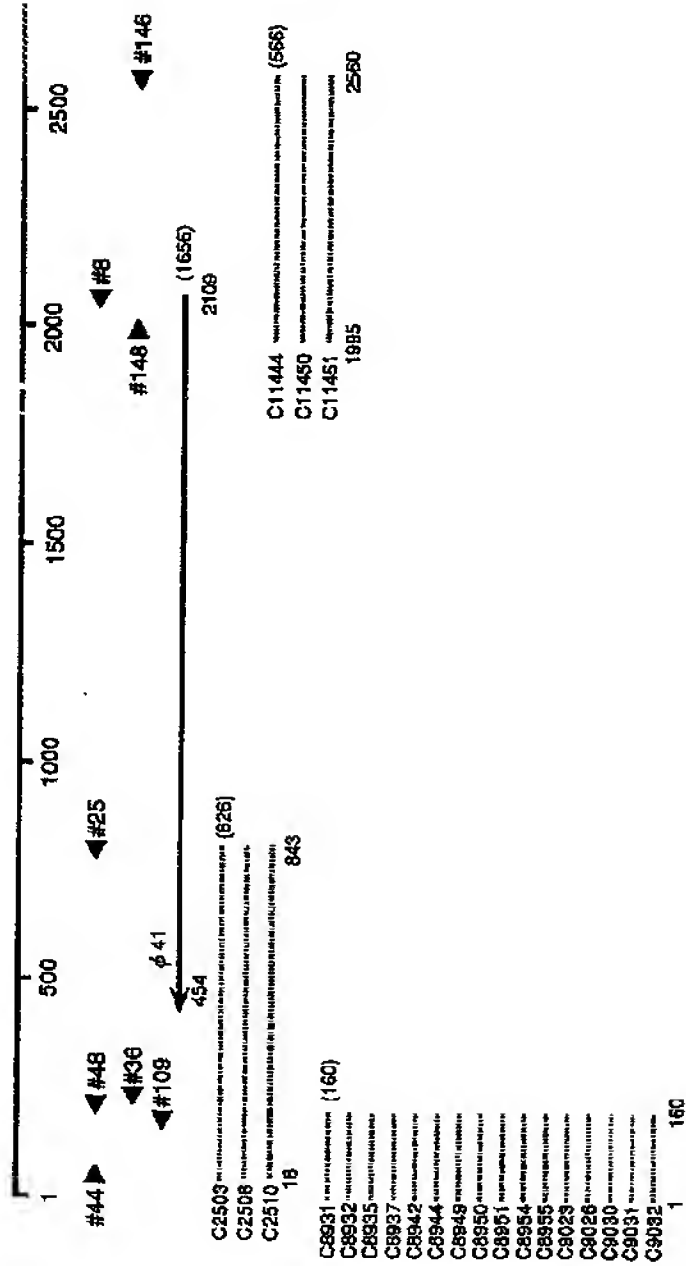
【書類名】 図面

【図 1】



【図 1】

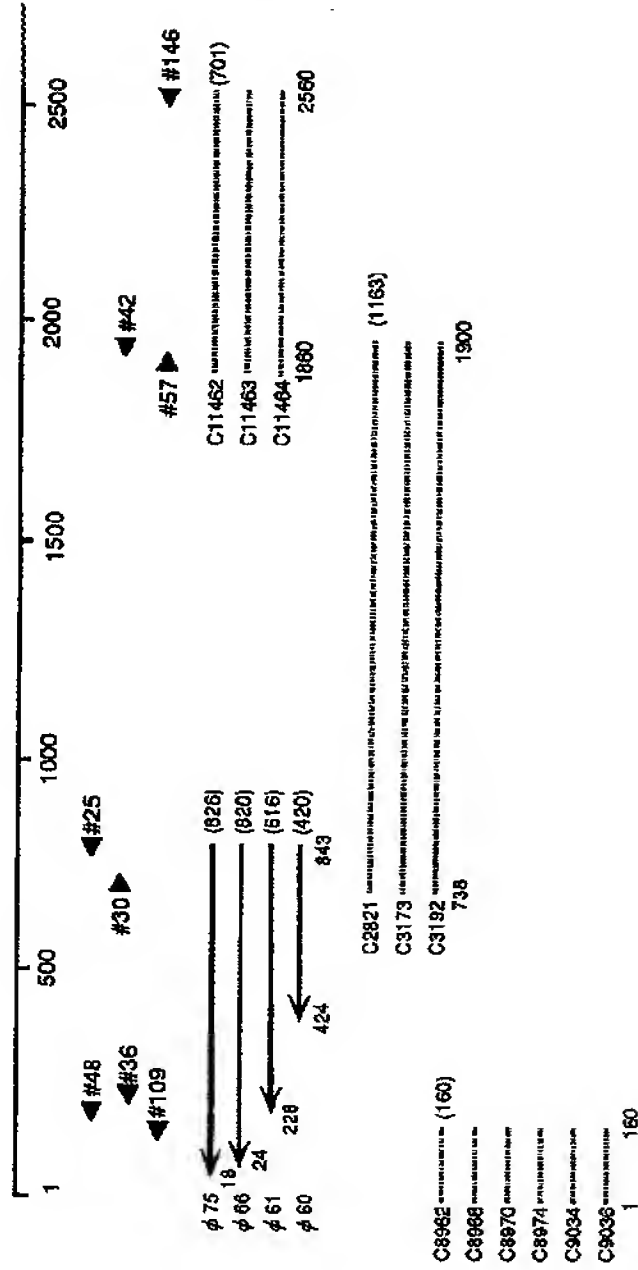
【 図 2 】



【 図 2 】

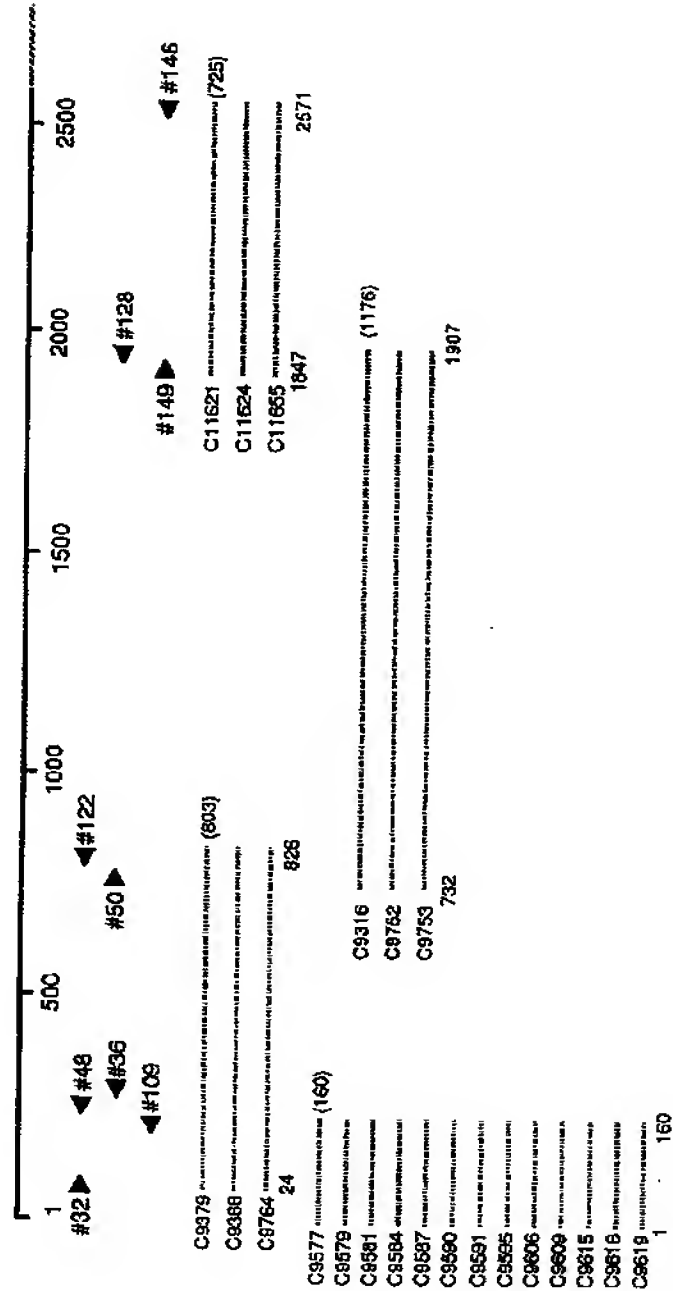
【 図 3 】

【 図 3 】

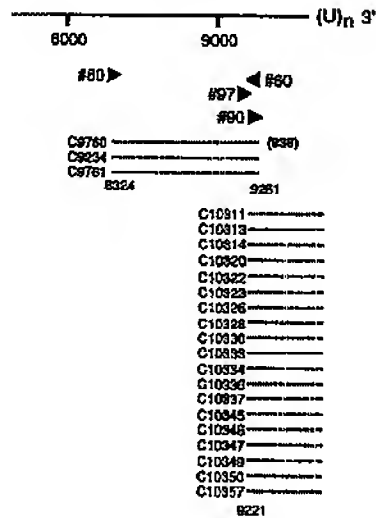


【 図 4 】

【 図 4 】



【図 5】



【手続補正書】

【提出日】平成 5 年 8 月 1 1 日

【手続補正 1】

【補正対象書類名】明細書

【補正対象項目名】発明の詳細な説明

【補正方法】変更

【補正内容】

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、非 A 非 B 型肝炎ウイルス（以下「NANB 型肝炎ウイルス」と略記する）遺伝子、これに関連するポリヌクレオチド、ポリペプチド、ならびに関連抗原、抗体検出系に関する。

【0002】

【従来の技術】ウイルスに由来する肝炎には、既に病原体が解明されその診断法や予防法が確立したものとして A 型肝炎、B 型肝炎がある。これ以外は一括して、非 A 非 B 型肝炎（以下「NANB 肝炎」という）と呼ばれてきた。日本における輸血後肝炎は、B 型肝炎の検出法が輸血血液のスクリーニングに導入された後は格段に減少したが、なお、年間推定 2 8 万例が、NANB 型肝炎ウイルスにより発症していると考えられてきた。

【0003】近年、NANB 型肝炎ウイルス肝炎は、タイプ別に C 型、D 型、E 型と命名され、その病原体についての研究が進展し、撲滅に向けての努力が世界的に進められるようになった。輸血後肝炎に関して、1988 年アメリカのカイロン社が C 型肝炎ウイルス（以下「HCV」と略記する）の RNA ウイルスゲノムのクローン化に成功したと発表し、これを基に HCV 抗体の測定系を

開発した。現在、HCV 抗体測定系は、輸血血液のスクリーニングや肝疾患患者の診断に、日本を始めとして各国で用いられている。この HCV 抗体の測定系は NANB 肝炎との関連性が確かに一部では認められる。しかし、キャリアや慢性肝炎の捕捉率は約 70% にすぎず、また急性期の抗体検出ができない等重大な問題点が残されている。したがって、カイロン社の前記開発によっても NANB 肝炎への対処は依然として解決されていない。NANB 肝炎の経過は不良であり、水平感染によっても持続感染化（キャリア化）し、慢性肝炎に進展することが確認されている。また、肝硬変、肝がんへ進展するケースも多いと予想され、一日も早いウイルス本体の全貌解明が望まれていた。

【0004】

【発明が解決しようとする課題】上記カイロン社の HCV 抗体検出キットでは診断できない NANB 肝炎が多数存在することから、日本における NANB 型肝炎ウイルスゲノムがカイロン社の解明した HCV ゲノムとは必ずしも一致せず、サブタイプの違いがあると推定された。したがって、特異性、感度においてさらに優れた NANB 肝炎診断薬、および有効なワクチンを開発するためには、サブタイプ別に NANB 型肝炎ウイルスの遺伝子レベル、アミノ酸レベルでの解析を完成することが望まれていた。本発明の目的は、NANB 型肝炎ウイルスゲノムの塩基配列を、全体にわたって完全に解明し、これに基づいてアミノ酸レベルでの解析を行うこと、ならびにこれを用いて NANB 型肝炎ウイルス関連抗原、抗体の検出系の確立を図ることである。

【0005】

【課題を解決するための手段】本発明者らは、NANB肝炎ウイルスの治療、予防、検査に不可欠なポリヌクレオチドを見出すために鋭意研究をつづけ、その結果、ヒトおよびチンパンジーキャリアの血清よりNANB肝炎ウイルスのRNAを単離し、非コード領域を含む遺伝子領域全体をカバーするcDNAをクローン化してその全部の塩基配列を決定し、さらにアミノ酸配列を決定した。その結果本発明者らは、NANB肝炎ウイルスであるHC-J6ゲノムの遺伝子全体の塩基配列を完全に解明することに成功した。かかる遺伝子及びアミノ酸レベルでの研究によって本発明は完成された。本発明のNANB肝炎ウイルス(HC-J6ゲノム)は、カイロン社のHCVゲノムとは別異のものである。

【0006】本発明は、配列番号1記載の塩基配列を有する非A非B型肝炎ウイルス遺伝子の発明であり、RNA 5'末端より340箇のヌクレオチドからなる非コード領域、これに続く構造蛋白質および非構造蛋白質をコードする領域を含む9099箇のヌクレオチド、ならびに3'末端108箇のUストッチを含む150箇のヌクレオチドからなる、配列番号1記載のウイルス遺伝子の発明であり、これらと、部分的置換を有する非A非B型肝炎ウイルス変異株遺伝子の発明である。また、配列番号2記載の塩基配列を有するポリヌクレオチドN-9589、配列番号3記載の塩基配列を有するcDNAクローンJ6-φ81、配列番号4記載の塩基配列を有するcDNAクローンJ6-φ8、これら各ヌクレオチドと部分的置換を有する非A非B型肝炎ウイルス変異株ヌクレオチドの発明である。さらに前記遺伝子またはポリヌクレオチドにコードされるポリペプチド、配列番号5記載のアミノ酸配列を有するポリペプチドP-3033、前記遺伝子、ポリヌクレオチド、cDNAクローンの全部または一部を宿主細胞に組み込み発現させて得たポリペプチド、前記各ポリペプチドに対するポリクローナル抗体またはモノクローナル抗体の発明である。さらに、前記各ポリペプチドを用いた非A非B型肝炎検出系、前記各抗体を用いた非A非B型肝炎検出系の発明である。

【0007】本発明のNANB肝炎ウイルス遺伝子は次のようにして得て、その塩基配列を決定した。ヒトおよびチンパンジー血漿より得た検体、HC-J1、HC-J4、HC-J6を用いて本研究を進めた。HC-J1、HC-J6は日本人供血者由来で、いずれもHCV抗体陽性と判定された検体より得た。HC-J4はNANB肝炎の感染性を確認したチンパンジーから得た検体に由来するが、しかしこれはカイロン社の前記のHCV抗体では(一)であった。この3種の血漿検体よりRNAを調整し、これらによってRNAの塩基配列を調べた。このうちRNAの5'末端側の約2500、および3'末端側の約1100の塩基配列については既に本年4

月26日付特許出願(特願平3-191376号)ならびに6月20日付特許出願においてその一部を開示した。本発明者らは、さらに研究を進め、今回HC-J6について、新たに非構造蛋白質をコードする領域のcDNAクローンの構造も解明し、これによって、HC-J6ゲノムの完全長RNA9589塩基の配列の研究を完成させ、本発明を完成させるに至った。

【0008】HC-J6ゲノムは、実施例に示すように、5'末端より340箇のヌクレオチドよりなる5'非コード領域を有し、それに続いて構造蛋白質をコードする領域が続き、さらに非構造蛋白質をコードする領域を有する。他方、3'末端側は、非構造蛋白質をコードする領域に続いて108塩基のUストッチを含む150塩基の非コード領域を有する。5'末端側から341番目のアデニンより始まるコード領域は9099塩基の一本の長いオープンリーディングフレームよりなり、これが3033アミノ酸をコードしていることが明らかとなった。HCV、すなわちC型肝炎ウイルスはその遺伝子構造からフラビウイルスに近縁であると考えられているが、本発明のNANB肝炎ウイルス遺伝子のコード領域も図1に示す如く、C(コア)、E(エンベロープ)、NS-1(非構造蛋白-1)、NS-2(同-2)、NS-3(同-3)、NS-4(同-4)、NS-5(同-5)からなると考えられる。

【0009】本発明のHC-J6ゲノムは、ヨーロッパ特許公開第388、232号(カイロン社)に示されたHCVゲノムと比較すると、ヌクレオチド配列について全領域では67.9%の相同性を、またアミノ酸配列については全領域で72.3%の相同性を示すにすぎない。領域ごとに個別に相同性を調べると、5'末端非コード領域において塩基配列で94.4%、コア領域においてアミノ酸配列で90.1%であり、これらの領域では比較的高い相同性を示している。他方、エンベロープ部分より下流の領域ではE、NS-1、NS-2、NS-3、NS-4、NS-5のそれぞれのアミノ酸配列で60.4%、71.1%、57.8%、81.1%、73.1%、69.9%、3'末端非コード領域の塩基配列で29.6%といずれも低い相同性を示すにすぎず、結局、HC-J6株がカイロン社の発見したHCV株と大きく異なることが明らかとなった。

【0010】本発明者ら既に公表した(Japan. J. Exp. Med., Vol. 60, 3, 167-177)HC-J1(アメリカ型)、HC-J4(日本型)株とのアミノ酸配列の比較では、コア領域では90%以上の相同性を示すのに対して、エンベロープ領域ではそれぞれ60.9%、53.1%と低い相同性を示すにすぎなかった。本発明により、HC-J6株は、HC-J1株、HC-J4株のいずれとも異なる別箇のタイプのウイルスであることが明らかとなった。

【0011】

【作用】本発明の遺伝子、ポリヌクレオチド、cDNAクローンは、その全部または一部を大腸菌、枯草菌等の宿主細胞に組み込んでポリペプチドを発現させることができる。本発明のポリペプチドは、NANB型肝炎の検出系として用いることができ、さらに、これに対するポリクローナル抗体またはモノクローナル抗体を製造することができる。本発明のポリクローナル抗体またはモノクローナル抗体は、非A非B型肝炎検出系として用いることができる。

【0012】

【実施例】以下、本発明の実施例について述べるが、もとより本発明がこれらの実施例に限定されるものではない。

【0013】 実施例1

(1) RNA抽出

日本人献血者の血漿から得た、HCV抗体陽性（オース・ダイアグノスティック・システムズ社製、オースHCV Ab ELISAキット）と判定された検体HC-J1、HC-J6およびNANB型肝炎の感染を確認したチンパンジーから得たHCV抗体（-）の検体HC-J4より、次のようにしてRNAを抽出した。血漿にトリス塩酸緩衝液（10mM、pH8.0）を加え、68×10³ rpmで1時間遠心した。得られたペレットに200mM NaCl、10mM EDTA、2%（w/v）ドデシル硫酸ナトリウム（SDS）、および1mg/mlのプロテナーゼKを含むトリス塩酸緩衝液（50mM、pH8.0）を加え、60℃で1時間加温し、エタノール沈澱を行い、RNAを得た。

【0014】 (2) HC-J1 cDNA合成

HC-J1血漿より抽出したRNAを70℃で1分間加温し、これを鋳型として10ユニットの逆転写酵素（cDNA Synthesis System Plus、アマシャム社）およびオリゴヌクレオチドプライマー（20-mer）20pmolを加えて、42℃、15時間反応させてcDNAを得た。プライマーとしては、ヨーロッパ特許出願公開318、216号に示されたHCV塩基配列を参照して合成した#8（5'-GATGCTTGCGGAACGCAATCA-3'）を用いた。

【0015】 (3) ポリメラーゼチェーンリアクション

nt 24~826

#32 (5'-ACTCCACCATAGATCACTCC-3')

#122 (5'-AGGTTCCCTGTTGCATAATT-3')

クローン: C9397、C9388、C9764

nt 732~1907

#50 (5'-GCCGACCTCATGGGGTACAT-3')

#128 (5'-TCGGTCGTGCCCACTACCAC-3')

クローン: C9316、C9752、C9753

nt 1867~2591

#149 (5'-TCTGTGTGTGGCCCACTGTA-3')

(PCR) によるcDNAの増幅

Gene Amp DNA増幅試薬キット（パーキンエルマー・シータス社）を用いたSaikiらの方法（Science 239, 487-491, 1988）により、DNAサーマイクラー（パーキンエルマー・シータス社）にて35サイクルのcDNA増幅を行った。

【0016】 (4) cDNAライブラリーの構築によるHC-J1、HC-J4の5'末端側塩基配列の決定

cDNAライブラリーを用いたHC-J1ゲノムの5'末端側の塩基配列解析は図2（HC-J1）図3（HC-J4）に示すように、cDNAをバクテリオファージλgt10に挿入して得たクローンの解析及びcDNAをPCRにて増幅して得られたクローンの解析結果の両者を併せて決定した。図2および図3はNANB肝炎ウイルスゲノムの5'末端を制限酵素切断部位とともに示し、使用するプライマーの位置も示す。図中、実線はバクテリオファージλgt10のライブラリーによるクローンで塩基配列を決定した範囲を、点線はPCRによるクローンで塩基配列を決定した範囲を示す。HC-J1のnt 454~2109の1656塩基はプライマー#8から得たcDNAをλgt10ファージベクター（アマシャム社）に挿入して得られたクローンφ41により決定した。つぎにこのシーケンスをもとに合成した、nt 824~843の新しいプライマー#25（5'-TCCCTGTTGCATAGTTCAAG-3'）を用いて、HC-J4のcDNAライブラリーから順次4つのcDNAクローンφ60、φ61、φ66、φ75を得て、これらのコンセンサスシーケンスから上流のnt 18~843の塩基配列を決めた。

【0017】 (5) HC-J6、5'末端側塩基配列の決定

HC-J6ゲノムの5'末端側の塩基配列は図4に示すように、cDNAをPCRにて増幅して得られたクローンの解析結果により決定した。HC-J6検体について、抽出したRNAから前記(2)の方法に従って、cDNAを抽出し、そのシーケンスを決定した。nt 24~2551はHC-J4の塩基配列を基にしたプライマーの組み合わせによるPCRにて得られた各クローンのコンセンサスシーケンスから決められた。

#146 (5' -AGTAGCATCATCCACAAGCA-3')

クローン: C11621、C11624、C11655

さらに上流の5'末端側を特定するため、nt246~265のアンチセンスプライマー#36 (5' -AACACTACTCGGCTAGCAGT-3') を用いてcDNAを合成したのち、ターミナルデオキシヌクレオチジルトランスフェラーゼによりcDNAの5'末端にdA付加を行い、2段階のone-sided PCR増幅を行った。すなわち、1回目はオリゴdTプライマー (20-mer) とnt188~207のアンチセンスプライマー#48 (5' -GTTGATCCAAGAAAGGACCC-3') を用いて35サイクルのPCR増幅を行い、2回目はそのPCR産物を鋳型にしてオリゴdTプライマー (20-mer) とnt140~160のアンチセンスプライマー#109 (21-mer; 5' -ACCGGATCCGCAGACCACTAT-3') を用いて30サイクルのPCRを行った。得られたPCR産物をM13ファージベクターにサブクローニングした。完全長の5'末端配列を有すると考えられる13個の独立したクローン、C9577、C9579、C9581、C9584、C9587、C9590、C9591、C9595、C9606、C9609、C9615、C9616、C9619を得、それらのコンセンサスシーケンスからnt1~23の塩基配列を決めた。

(6) HC-J6、中央領域塩基配列の決定

HC-J6血漿100mlを原料に、(2)の方法に準じてλgt10バクテリオファージにcDNAライブラリーを構築した。プライマーとして、ヨーロッパ特許公開318、216号に示されたHCV塩基配列を参照して合成したポリヌクレオチド#162および#81を用い、クローンの選択はブラークハイブリダイゼーションにより行った。図1に示すように、得られた4個のcDNAクローンφ2 (nt6996~8700)、φ6 (nt6485~8700)、φ8 (nt6008~8700)、φ81 (nt2199~6168) を解析し、得られたコンセンサスシーケンスからnt2552~8700の塩基配列を決定した。このうち、クローンφ81、φ8はそれぞれ配列番号3、4の塩基配列を有する。

【0018】(7) HC-J6ゲノムの3'末端側の塩基配列は図5に示すように、PCRにて増幅して得られたクローンの解析結果より決定した。HC-J6のnt8701~9241の配列については検体をプライマー#80 (5' -GACACCCGCTGTTTTGACTC-3') および#60 (5' -GTTCTTACTGCCAGTTGAA-3') を用いたPCRにかけ、得られた938塩基のクローンC9760、C9234、C9761の塩基配列より得られるコンセンサスシーケンスから決定した。nt9242より下流の

3'末端側塩基配列は以下の方法の従って決定した。すなわち、各検体より(1)記載の方法に従ってRNAを抽出し、ポリ(A)ポリメラーゼを用いてRNAの3'末端にポリ(A)を付加し、オリゴ(dT)20をプライマーとしてcDNA合成を行い、得られたcDNAをテンプレートとしてPCRに供した。PCRは第1段階としてセンスプライマーとして#97 (5' -AGTCAGGGCGTCCCTCATCT-3') を、アンチセンスプライマーとしてオリゴ(dT)20を用いて行った。次に、得られたcDNAを第2段階PCRとして第1段階より下流に相当するセンスプライマー#90 (5' -GCCGTTTGCGGCCGATATCT-3') を、アンチセンスプライマーとしてオリゴ(dT)20を用いて行った。2段階のPCRで得られた増幅産物にT4DNAポリラーゼを作用させ両端を平滑にした後、T4ポリヌクレオチドキナーゼにて5'末端をリン酸化し、M13mp19ファージベクターのHincII部位にサブクローニングして塩基配列を決定した。こうして得られた19個のクローン、C10311、C10313、C10314、C10320、C10322、C10323、C10326、C10328、C10330、C10333、C10334、C10336、C10337、C10345、C10346、C10347、C10349、C10350、C10357から得られるコンセンサスシーケンスより3'末端側の塩基配列を決定した。以上の解析結果よりHC-J6ゲノムに相当するcDNA塩基配列を配列番号2記載のように決定し、またゲノムの完全長RNA塩基配列を配列番号1記載のように決定した。

【0019】(8) HC-J6、アミノ酸配列の決定
HC-J6ゲノムのRNA塩基配列をもとに、開始コドンATGより始まるコード領域のアミノ酸配列を配列番号5のように決定した。その結果、HC-J6は3033アミノ酸残基からなるポリペプチド前駆体をコードする1本の長いオープンリーディングフレームよりなることが明らかになった。

【0020】

【発明の効果】本発明は、非A非B型肝炎ウイルスHC-J6を単離し、そのウイルス遺伝子の全塩基配列を解明したことにもとづく発明であって、非A非B型肝炎ウイルスを高感度に検出することができ、また、これに必要なポリペプチド、ポリクローナル抗体、モノクローナル抗体を提供することができる。

【0021】

【配列表】

配列番号: 1

配列の長さ: 9589

配列の型: 核酸

鎖の数：一本鎖

トポロジー：直鎖状

配列の種類：genomic RNA

特徴を決定した方法：E

配列

```

ACCCGCCCCU AAUAGGGGCG ACACUCCGCC AUGAACCACU CCCUGUGAG GAACUACUGU 60
CUUCACGCAG AAAGCGUCUA GCCAUGGCGU UAGUAUGAGU GUCGUACAGC CUCCAGGCC 120
CCCCUCCCCG GGAGAGCCAU AGUGGUCUGC GGAACCGGUG AGUACACCGG AAUUGCCGGG 180
AAGACUGGGU CCUUUCUUGG AUAAACCCAC UCUAUGCCCG GUCAUUUGGG CGUGCCCCCG 240
CAAGACUGCU AGCCGAGUAG CGUUGGGUUG CGAAAGGCCU UGUGGUACUG CCUGAUAGGG 300
UGCUGCGAG UGCCCCGGGA GGUCUCGUAG ACCGUGCACC AUGAGCACA AUCCUAAACC 360
UCAAAGAAAA ACCAAAAGAA ACACCAACCG UGCCCCACAA GACGUUAAGU UUCCGGGCGG 420
CGGCCAGAUU GUUGGGGGAG UAUACUUGUU GCCGCGCAGG GGCCCCAGGU UGGGUGUGCG 480
CGCGACAAGG AAGACUUCGG AGCGGUCCCA GCCACGUGGA AGGCGCCAGC CCAUCCCUAA 540
GGAUCGGCGC UCCACUGGCA AAUCCUGGGG AAAACCAGGA UACCCUGGC CCCUAUACGG 600
GAAUGAGGGA CUCGGCUGGG CAGGAUGGCU CCUGUCCCC CGAGGUUCCC GUCCUCUUG 660
GGGCCCCAAU GACCCCCGGC AUAGGUCCCG CAACGUGGGU AAGGUCAUCG AUACCCUAAC 720
GUGCGGCUUU GCCGACCUCU UGGGUUACAU CCCUGUCGUA GCGCCCCCGC UCGGCGGCGU 780
CGCCAGAGCU CUCGCGCAUG GCGUGAGAGU CCUGGAGGAC GGGGUUAAU UUGCAACAGG 840
GAACUUAACC GGUUGCUCCU UUUCUAUCUU CUUGCUGGCC CUGCUGUCCU GCAUACACC 900
CCCGGUCUCC GCUGCCGAAG UGAAGAACAU CAGUACCGGC UACAUGGUGA CCAACGACUG 960
CACCAUAGAU AGCAUUACCU GGCAACUCCA GGCUGCUGUC CUGGACGUCC CCGGGUGCGU 1020
CCCGUGCGAG AAAGUGGGGA AUACAUCUCG GUGCUGGAUA CCGGUCUCAC CGAAUGUGGC 1080

```

CGUGCAGCAG CCCGGCGCCC UCACGCAGGG CUUACGGACG CACAUUGACA UGGUUGUGAU 1140
 GUCCGCCACG CUCUGCUCGG CUCUUUACGU GGGGGACCUC UGGGGUGGGG UGAUGCUUGC 1200
 AGCCAGAUUG UUCAUUGUCU CGCCACAGCA CCACUGGUUU GUGCAAGACU GCAAUUGCUC 1260
 CAUCUACCCU GGUACCAUCA CUGGACACCG CAUGGCGUGG GACAUGAUGA UGAACUGGUC 1320
 GCCCACGGCU ACCAUGAUCC UGGCGUACGC GAUGCGCGUC CCCGAGGUGA UCAUAGACAU 1380
 CAUUGGCGGG GCUCAUUGGG GCGUCAUGUU CGGCUUAGCC UACUUCUCUA UGCAGGGAGC 1440
 GUGGGCAAAA GUCGUUGUCA UUCUUUUGCU GGGCGCCGGG GUGGACGCGC AAACCCAUAC 1500
 CGUUGGGGU UGUACCGCGC AUAACGCCAG GACCCUACCC GGCAGUUCU CCCUUGGUGC 1560
 CAGGCAGAAA AUCCAGCUCU UCAACACCAA UGCGAGUUGG CACAUCAACC GCACCGCCCU 1620
 GAACUGCAAU GACUCUUGC ACACCGGCUU CCUCGCGUCA CUGUUCUACA CCCACAGCUU 1680
 CAACUCGUCA GGAUGUCCCG AACGCAUGUC CGCCUGCCGC AGUAUCGAGG CCUUUCGGGU 1740
 GGGUUGGGG GCUUUAACAU AUGAGGACAA UGUACCAAU CCAGAGGAUA UGAGACCGUA 1800
 UUGCUGGCAC UACCCACCAA GACAGUGUGG UGUAGUCUCC GCGAGCUCUG UGUGUGGCC 1860
 AGUGUACUGU UUCACCCCCA GCGCAGUAGU AGUGGUUACG ACCGAUAGAC UUGGAGCGCC 1920
 CACUACACG UGGGGGGAGA AUGAGACAGA UGUUUCUUA UUGAACAGCA CUCGACCACC 1980
 GCAGGGGUGA UGGUUCGGCU GCACGUGGAU GAACUCCACU GGCUACACCA AGACUUGCGG 2040
 CGCACACCC UGCCGCAUUA GAGCUGACUU CAAUGCCAGC AUGGACUUGU UGUGCCCCAC 2100
 GGACUGUUUU AGGAAGCAUC CUGAUACCAC CUACAUCAAA UGUGGCUCUG GGGCCUGGCU 2160
 CACGCCAAGG UGCCUGAUGG ACUACCCCUA CAGGCUCUGG CAUUAACCCU GCACAGUUA 2220
 CUAUACCAUC UUCAAAUUA GGAUGUUGU GGGGGGGGUC GAGCACAGGC UCACGGCUGC 2280
 GUGCAAUUUC ACUGGUGGGG AUCGUUGCAA CUUGGAGGAC AGAGACAGAA GUCAACUGUC 2340
 UCCUUGGUG CACUCCACCA CGGAGUGGGC CAUUUUACCU UGCACUACU CGGACCUGCC 2400
 CGCCUUGUG ACUGGUCUUC UCCACCUCU CCAAACAUC GUGGACGUGC AAUUCAGUA 2460
 UGGCUAUCA CCUGCUCUCA CAAAUAACU CGUCCGAUGG GAGUGGGUAG UACUCUUAU 2520
 CCUGCUCUA GCGGACGCCA GGGUUUGCGC CUGCUUAUGG AUGCUAUCU UGUUGGGCCA 2580
 GCGCGAAGCA GCACUAGAGA AGUUGGUCGU CUUGCACGCU GCGAGCGCAG CUAGCUGCAA 2640
 UGGCUUCCUA UACUUUGUCA UCUUUUUGCU GGCUGCUUGG UACAUCAGG GUCGGGUAGU 2700
 CCCUUGGCU ACUUAUUCU UCACUGGCCU AUGGUCCUUU GGCUACUGC UCCUAGCAU 2760
 GCGCAACAG GCUUAUGCUU AUGACGCAUC UGUACAUGU CAGAUAGGAG CAGCUCUGU 2820

GGUACUGAUC ACUCUCUUA CACUCACCCC CGGUAUAAG ACCCUUCUA GCGGCUUUCU 2880
 GUGGUGGUUG UGCUAUCUUC UGACCCUGGC GGAAGCUAUG GUCCAGGAGU GGGCACCACC 2940
 UAUHCAGGUG CGCGGUGGCC GUGAUGGGAU CAUAUGGGCC GUCCCAUAU UCUGCCCGGG 3000
 UGUGGUGUUU GACAUAAACA AGUGGCUCUU GCGGUGCUU GGGCCUGCUU AUCUCCUAAA 3060
 AGGUGCUUUG ACGCGUGUGC CGUACUUCGU CAGGGCUCAC GCUCUACUAA GGAUGUGCAC 3120
 CAUGGUAAGG CAUCUCGCGG GGGGUAGGUA CGUCCAGAUG GUGCUACUAG CCCUUGGCAG 3180
 GUGGACUGGC ACUUAACUUC AUGACCACCU CACCCCUAUG UCGGAUUGGG CUGCUAAUGG 3240
 CCUGCGGGAC UUGGCGGUCG CCGUGGAGCC UAUCAUCUUC AGUCCGAUGG AGAAAAAGU 3300
 CAUCGUCUGG GGAGCGGAGA CAGCUGCUUG GGGGGAUUC UUACACGGAC UGCCCGUGUC 3360
 CGCCCGACUU GCGCGGAGG UCCUCCUUGG CCCAGCUGAU GGCUAUACCU CCAAGGGGUG 3420
 GAGUCUUCUC GCGCCCAUCA CUGCUUAUGC CCAGCAGACA CGCGGCCUUU UGGGCACCAU 3480
 AGUGGUGAGC AUGACGGGGC GCGACAAGAC AGAACAGGCC GGGGAGAUUC AGGUCCUGUC 3540
 CACGGUCACU CAGUCCUUC UCGGAACAAC CAUCUCGGGG GUCUUAUGGA CUGUCUACCA 3600
 UGGAGCUGGC AACAAGACUC UAGCCGGCUC ACGGGGUCG GUCACACAGA UGUACUCCAG 3660
 UGCUGAGGGG GACUUAUGG GUGGGCCAG CCCCCCGGG ACCAAAUUUU UGGAGCCGUG 3720
 CACGUGUGGA GCGGUCGACC UAUACCUGGU CACGCGAAAC GCUGAUGUCA UCCCGGCUCG 3780
 AAGACCGGGG GACAAGCGAG GAGCGUACU CUCCCCGAGA CCUCUUCCA CUUGAAGGG 3840
 GUCCUCGGGG GCGCGGUGC UCUGCCCCAG AGGCCACGCU GUCGGGGUCU UCCGGGCAGC 3900
 CGUGUGCUCC CGGGGCGUGG CCAAGUCCA AGAUUUUAUC CCCGUGAGA CACUUGACAU 3960
 CGUACUCGG UCCCCACCU UAAGUGACAA CAGCACACCA CCUGCUGUGC CCCAAACUUA 4020
 UCAGGUCGGG UACUUAUAG CCCCAGUGG UAGUGGAAAG AGCACCAGG UCCUGUCGC 4080
 GUAUGCCGCU CAGGGGUACA AAGUGCUAGU GCUUAUCCC UCGGUGGCUG CCACCCUGGG 4140
 GUUUGGGGCG UACUUGUCCA AGGCACUUG CAUCAAUCCC AACAUUAGGA CUGGGGUCAG 4200
 GACUGUGACG ACCGGGCGCG CCAUCACGUA CUCCACAUU GCAAAAUUCC UCGCCGAUGG 4260
 GGGCUGCGCA GCGGCGCCU AUGACAUCU CAUAUGCGAU GAAUGCCAUG CCGUGGACUC 4320
 UACCACCAUU CUCGGCAUCG GAACAGUCCU CGAUCAGCA GAGACAGCCG GGGUCAGGCU 4380
 AACUGUACUG GCUACGGCUA CGCCCCCGG GUCAGUGACA ACCCCCACC CCAACAUAGA 4440
 GGAGGUGGCC CUCGGGCAGG AGGGUGAGAU CCCCUUCUUA GGGAGGGCGA UGCCCGUGUC 4500
 AUACAUCAAG GGAGGAAGAC ACUUGAUCUU CUGCCACUCA AAAAAAGU GUGACGAGCU 4560

CGCGGCGGCC CUUCGGGGUA UGGGCUUGAA CGCAGUGGCA UACUACAGAG GGCUGGACGU 4620
 CUCCGUAAUA CCAACUCAGG GAGACGUAGU GGUCGUCGCC ACCGACGCCC UCAUGACGGG 4680
 GUUUACUGGA GACUUUGACU CCGUGAUCGA CUGCAACGUA GCGGUCACUC AAGUUGUAGA 4740
 CUUCAGCUUG GACCCACAU UCACCAUAAC CACACAGACU GUCCUCAAAG ACGCUGUCUC 4800
 ACGUAGCCAG CGCGGGGGCC GCACGGGCAG GGAAGACUG GGUUUUUAUA GGUUUGUUUC 4860
 CACUGGUGAG CGAGCCUCAG GAAUGUUUGA CAGUGUAGUG CUCUGCGAGU GCUACGAUGC 4920
 AGGGGCCGCA UGGUUGGAGC UCACACCAGC GGAGACCACC GUCAGGCUCA GAGCAUAUUU 4980
 CAACACACCU GGUUUGCCUG UGUCCAAGA CCAUCUUGAG UUUUGGGAGG CAGUUUUCAC 5040
 CGGCCUCACA CACAUAGAUG CCCACUCCU UCCCCAACA AAGCAAUCGG GGGAAAAUUU 5100
 CGCAUACUUA ACAGCCUACC AGGCUACAGU GUGCGCUAGG GCCAAAGCCC CCCCCCGUC 5160
 CUGGGACGUC AUGUGGAAGU GUUUGACUCG ACUCAAGCCC ACACUCGUGG GCCCCACACC 5220
 UCUCUGUAC CGCUUGGGCU CUGUUACCA CGAGGUCACC CUCACGCAUC CUGUGACGAA 5280
 AUACAUCCGC ACCUGCAUGC AAGCCGACCU UGAGGUAUG ACCAGCACGU GGUUCUAGC 5340
 UGGGGGGGUC UUGGGGGCCG UCGCCCGGUA CUGCCUGGG ACCGGGUGUG UUGCAUCAU 5400
 CGGCCGCUUG CACGUUAACC AGCGAGCCGU CGUUGCACC GACAAGGAGG UCCUCUAUGA 5460
 GGCUUUUGAU GAGUUGGAGG AAUGUGCCUC UAGAGCGGCU CUCAUUGAAG AGGGGACGG 5520
 GAUAGCCGAG AUGCUGAAGU CCAAGAUCCA AGGCUUAUUG CAGCAAGCUU CCAACAAGC 5580
 UCAAGACAU CAACCCGCUG UGCAGGCUUC UUGGCCAAG GUAGAGCAU UCHGGGCCAA 5640
 ACACAUUGG AACUUCAUCA GCGGCAUUA AUACCUCGCA GGACUAUCAA CACUGCCAGG 5700
 GAACCCUGCU GUAGCUUCCA UGAUGGCAU CAGUGCCGCC CUCACCAGUC CGUUGUCAAC 5760
 UAGCACCACU AUCCUUCUA ACAUUUUGG GGGCUGGCUA GCAUCCAAA UUGCGCCUCC 5820
 CGCGGGGGCU ACCGGCUUCG UCGUCAGUGG CCUGGUGGG GCVGCCGUAG GCAGCAUAGG 5880
 CUUGGGUAAG GUGCUGGUGG ACAUCCUGGC AGGGUAUGGU GCGGGCAUUU CGGGGGCUCU 5940
 CGUCGCAUUC AAGAUAUGU CUGGCGAGAA GCCCUCAUG GAGGAUGUUG UCAACCUGCU 6000
 GCCUGGAUU CUGUCUCCGG GUGCCUGGU GGUGGGAGUC AUCUGGCGG CCAUCCUGCG 6060
 CGGACAGUG GGACCGGGGG AAGGCGGUGU CCAUUGGAG AAUAGGCUA UUGCCUUGC 6120
 UUCAGAGGA AACCAGUCG CCCCACCA CUACGUGAG GAGUCGAGU CGUCGCAGCG 6180
 UGUGACCAA CUACUUGGU CCCUACCAU AAGCAGCCUG CUCAGGAGAC UCCACAACUG 6240
 GAUUAUGAA GACUGCCCA UCCAUAGCAG CGGCUCGUG CUCGCGAUG UGUGGGAUUG 6300

GGUUUGCACC AUCCUAACAG ACUUUAAAA CUGGCUGACC UCCAAAUUGU UCCCAAAGAU 6360
 GCCUGGUCUC CCCUUUAUCU CUUGUCAAAA GGGGUACAAG GCGUGUGGG CUGGCACUGG 6420
 UAUCAUGACC ACACGGUGUC CUUGCGGCGC CAUAUUCUCU GGCAAUUUCG GCCUGGGUC 6480
 CAUGAGAAUU ACGGGGCCA AAACUGCAU GAAUAUCUGG CAGGGGACCU UUCCCAUCAA 6540
 UUGUACACG GAGGGCCAGU GCGUGCCGA ACCCGACCA AACUUUAAGA UCGCCAUCUG 6600
 GAGGGUGGCG GCCUCAGAGU ACGCGGAGGU GACGCAGCAC GGGUCAUACC ACUACAUAAC 6660
 AGGACUUACC ACUGAUAAU UGAAAGUUC UUGCCAACUA CCUUCUCCAG AGUUCUUUC 6720
 CUGGGUGGAC GGAGUGCAGA UCCAUAGGUU UGCCCCAU CCGAAGCCGU UUUUUCGGGA 6780
 UGAGGUCUGG UUCUGCGUG GGCUAAUUC AUUUGUCGUC GGGUCUAGC UCCCUUGCGA 6840
 UCCUGAACCU GACACAGACG UAUUGACGUC CAUGCUAACA GACCCAUCC AUUAACAGGC 6900
 GGAGACUGCA GCGCGGCUU UGGCAGGGG GUCACCCCG UCCGAGGCA GCUCCUAGC 6960
 GAGCCAGCUA UCGGCACCAU CGCUGCGAGC CACCUGCACC ACCCAGGCA AGGCCUAUGA 7020
 UGUGGACAU GUGGAUCCA ACCUGUUAU GCGGGGCGAU GUGACCCGA UAGAGUCUGA 7080
 GUCCAAAGUG GUCGUUCUG ACUCUCUGA CCCAUGGUC GAAGAAAGGA GCGACCUUGA 7140
 GCCUUGAUU CCAUCGGAU AUAUGCUC CAGAAGAGA UUCCACCAG CCUACCGGC 7200
 UUGGGACCG CCUGAUACA ACCCAGCGU UGUGGAUUG UGGAAGAGG CAGAUUACCA 7260
 ACCGGCACU GUUGCGGCU GCGUCUC CCCCCUAG AAAACCCGA GCGUCCCC 7320
 AAGGAGCGC CGGACAGUG GUCUGAGUGA GAGCUCAUA GCAGAUCCC UACAACAGCU 7380
 GGCAUCAAG UCCUUUGGC AGCCCCCC AAGCGCGAU UCAGGCCUU CCACGGGGC 7440
 GGACGAGCC GAUUCGGCA GUCGAGCGC CCGGAGAG UUGGCCUUU CGGAGACAG 7500
 UGCCAUUCC UCCAUGCCC CUCUGAGGG GAGCCUGGA GAUCCAGAC UGGAGCCUGA 7560
 GCAGGAGAG CUUCAACCUC CCCCCAGG GGGGUGGUA ACCCCGCGU CAGGCUCGG 7620
 GUCUUGGUC ACUUGUCCG AGGAGGAGA CUCCGUGUG UGCUGUCCA UGUCAUACUC 7680
 CUGGACCGG GCUCUAAUA CUCCUUGUAG CCGGAAGAG GAAAAGUUG CAAUUGGCC 7740
 CUUGAGCAAC UCCUGUUGG GAUAUCACA CAAGGUGUAC UGUACCACAU CAAAGAGCG 7800
 CUCAUUAAG GCUAAAAGG UAACUUUGA UAGGAUGCA GCGCUGGAC CUCAUUAUGA 7860
 CUCAGUCUG AAGGACAUU AGCUAGCGG CUCCAAGGUC ACCGCAAGG UUCUACUUU 7920
 AGAGGAGGC UGCAGUUA CUCCACCCA CUCUGCAAG UCCAAGUAG GGUUUGGGC 7980
 UAAGGAGGUC CCGAGCUUG CCGGAGAGC CUUAACCAC AUCAAGUCC UGUGGAAGGA 8040

```

CCUCCUGGAA GACACACAAA CACCAAUUCC UACAACCAUC AUGGCCAAAA AUGAGGUGUU 8100
CUGCGUGGAC CCCACCAAGG GGGGUAAGAA AGCAGCUCGC CUUAUCGUUU ACCCUGACCU 8160
CGGCGUCAGG GUCUGCGAGA AAAUGGCCCU UUAUGAUUUC ACACAAAAGC UUCCUCAGGC 8220
GGUGAUGGGG GCUUCUUAUG GAUUCAGUA CUCCCCGCU CAGCGGGUGG AGUUCUCUU 8280
GAAGGCAUGG GCGGAAAAGA AAGACCCUUA GGGUUUUUCG UAUGAUACCC GAUGCUUUGA 8340
CUCAACCGUC ACUGAGAGAG ACAUCAGGAC UGAGGAGUCC AUUAUUCGGG CUUGUCCUU 8400
GCCCAGGAG GCCCACACUG CCAUACACUC ACUGACUGAG AGACUUUACG UGGGAGGGCC 8460
CAUGUUC AAC AGCAAGGGCC AGACCUGCGG GUACAGGCGU UGCCGCGCCA GCGGGGUGCU 8520
UACCACUAGC AUGGGGAACA CCAUCACAU GUAUGUGAAA GCCUAGCGG CCUGUAAGGC 8580
UGCAGGGAUA AUUGCGCCCA CAAUGCUGGU AUGCGGCGAU GACUUGGUUG UCAUCUCAGA 8640
GAGCCAGGGG ACCGAGGAGG ACGAGCGGAA CCUGAGAGCC UUCACGGAGG CUAUGACCAG 8700
GUAUUCUGCC CCUCCUGGUG ACCCCCCAG ACCGGAUUAU GACCUGGAGC UGAUAACAUC 8760
UUGCUCCUCA AAUGUGUCUG UGGCGUUGG CCCACAAGGC CGCCGAGAU ACUACCUGAC 8820
CAGAGACCCU ACCACUCCAA UCGCCCGGGC UGCCUGGGA ACAGUUAGAC ACUCCCCUGU 8880
CAAUUCAGG CUAGGAAACA UCAUCCAGUA CGCCCCAACC AUUGGGCUC GCAUGGUCCU 8940
GAUGACACAC UUCUUCUCCA UUCUCAUGGC CCAAGAUACU CUGGACCAGA ACCUCAACUU 9000
UGAGAUGUAC GGAGCGGUGU ACUCCGUGAG UCCCUUGGAC CUCCAGCCA UAAUUGAAAG 9060
GUUACACGGG CUUGACGCUU UCUCUCUGCA CACAUACACU CCCCACGAAC UGACACGGGU 9120
GGCUACAGCC CUCAGAAAAC UUGGGGCGCC ACCCUCAGA GCGUGGAAGA GCCGGGCACG 9180
UGCAGUCAGG GCGUCCCUCA UCUCGUGG GGGGAGAGCG GCCGUUGCG GCCGAUAUCU 9240
CUUCAACUGG GCGGUGAAGA CCAAGCUCAA ACUCACUCCA UUGCCGGAAG CGCGCCUCCU 9300
GGAUUUAUCC AGCUGGUUCA CUGUCGGCGC CGCGGGGGC GACAUUUAUC ACAGCGUGUC 9360
GCGUGCCCGA CCCCUGUUAU UACUCCUUGG CCUACUCCUA CUUUUGUAG GGGUAGGCCU 9420
UUUCCUACUC CCCGUCGGU AGAGCGGCAC ACAUAGCUA CACUCCAUAG CUAACUGUCC 9480
CUUUUUUUUU UUUUUUUUUU UUUUUUUUUU UUUUUUUUUU UUUUUUUUUU UUUUUUUUUU 9540
UUUUUUUUUU UUUUUUUUUU UUUUUUUUUU UUUUUUUUUU UUUUUUUUUU 9589

```

【0022】配列番号：2

配列の長さ：9589

配列の型：核酸

鎖の数：一本鎖

トポロジー：直鎖状

配列の種類：cDNA to genomic RNA

特徴を決定した方法：E

配列

ACCCGCCCT AATAGGGGCG AACTCCGCC ATGAACCACT CCCCTGTGAG GAACTACTGT 60
 CTTACGCGAG AAAGCGTCTA GCCATGGCGT TAGTATGAGT GTCGTACAGC CTCAGGGCCC 120
 CCCCTCCCG GGAGAGCCAT AGTGGTCTGC GGAACCGGTG AGTACACCGG AATTGCCGGG 180
 AAGACTGGGT CCTTCTTGG ATAAACCCAC TCTATGCCCG GTCATTGGG CGTGCCCCCG 240
 CAAGACTGCT AGCCGAGTAG CGTTGGGTG CGAAAGGCCT TGTGGTACTG CCGATAGGG 300
 TGGTTGCGAG TGCCCCGGGA GGTCTCGTAG ACCGTGCACC ATGAGCACAA ATCCTAAACC 360
 TCAAAGAAAA ACCAAAAGAA ACACCAACCG TCGCCACAA GACGTTAAGT TTCCGGGCGG 420
 CGGCCAGATC GTTGGCGGAG TATACTTGTG GCCGCGCAGG GCGCCAGGT TGGGTGTGG 480
 CGCGACAAGG AAGACTTCGG AGCGGTCCCA GCCACGTGGA AGGCGCCAGC CCATCCCTAA 540
 GGATCGGCGC TCCACTGGCA AATCCTGGGG AAAACCAGGA TACCCCTGGC CCCTATACGG 600
 GAATGAGGGA CTCGGCTGGG CAGGATGGCT CCTGTCCCC CGAGGTTCCC GTCCCTCTG 660
 GGGCCCAAT GACCCCGGC ATAGGTCCCG CAACGTGGGT AAGGTATCG ATACCCTAAC 720
 GTGCGGCTTT GCGACCTCA TGGGTACAT CCTGTGTA GCGCCCCGC TCGGCGGCGT 780
 CGCCAGAGCT CTCGGCATG GCGTGAGAGT CCTGGAGGAC GGGGTAAAT TTGCAACAGG 840
 GAACTTACCC GGTGTCTCT TTTCTATCTT CTGTGTGGC CTGCTGTCT GCATACCAC 900
 CCGGTCTCC GCTGCCGAAG TGAAGAACAT CAGTACCGG TACATGGTGA CCAACGACTG 960
 CACCAATGAT AGCAATACCT GGCAACTCCA GGCTGTGTC CTCACGTCC CCGGGTGGT 1020
 CCCGTGCGAG AAAGTGGGA ATACATCTCG GTGCTGGATA CCGGTCTAC CGAATGTGG 1080
 CGTGCAGCAG CCGGGGCCC TCACGCAGGG CTTACGGAC CACATTGACA TGGTTGTGAT 1140

GTCCGCCACG CTCTGCTCCG CTCCTTACBT GGGGGACCTC TCCGGTGGGG TGATGCTTGC 1200
AGCCCAGATG TTCATTGTCT CGCCACAGCA CCACTGGITT GTGCAAGACT GCAATTGCTC 1260
CATCTACCCT GGTACCATCA CTGGACACCG CATGGCGTGG GACATGATGA TGAAGTGGTC 1320
GCCACGGCT ACCATGATCC TGGCGTACGC GATGCGCGTC CCCGAGGTCA TCATAGACAT 1380
CATTGGCGGG GCTCATTGGG GCGTCATGTT CGGCTTAGCC TACTTCTCTA TGCAGGGAGC 1440
GTGGGCAAAA GTCGTTGTCA TTCTTTTGCT GGGCGCGGGG GTGGACGGGC AAACCCATAC 1500
CGTTGGGGGT TCTACCGCGC ATAACGCCAG GACCCTCACC GGCATGTTCT CCCTTGGTGC 1560
CAGGCAGAAA ATCCAGCTCA TCAACACCAA TGGCAGTTGG CACATCAACC GCACCGCCCT 1620
GAACTGCAAT GACTCTTTGC ACACCGGCTT CCTCGCGTCA CTGTTCTACA CCCACAGCTT 1680
CAACTCGTCA GGATGTCCCG AACGCATGTC CGCCTGCCGC AGTATCGAGG CCTTTCGGGT 1740
GGGATGGGGC GCCTTACAAT ATGAGGACAA TGTACCAAT CCAGAGGATA TGAGACCGTA 1800
TTGCTGGCAC TACCCACCAA GACAGTGTGG TGTAGTCTCC GCGAGCTCTG TGTGTGGCCC 1860
AGTGTACTGT TTCACCCCCA GCCCAGTAGT AGTGGGTACG ACCGATAGAC TTGGAGCGCC 1920
CACTTACCG TGGGGGGAGA ATGAGACAGA TGTCTTCTA TTGAACAGCA CTCGACCACC 1980
GCAGGGGTCA TGGTTCGGCT GCACGTGGAT GAACTCCACT GGCTACACCA AGACTTGGCG 2040
CGCACCACCC TGGCGCATTG GAGCTGACTT CAATGCCAGC ATGGACTTGT TGTGCCCCAC 2100
GGACTGTTTT AGGAAGCATC CTGATACCAC CTACATCAAA TGTGGCTCTG GGGCCTGGCT 2160
CACGCCAAGG TGGCTGATCG ACTACCCCTA CAGGCTCTGG CATTACCCCT GCACAGTTAA 2220
CTATACCATC TTCAAAATAA GGATGTATGT GGGGGGGGTC GAGCACAGGC TCACGGCTGC 2280
GTGCAATTC ACTCGTGGGG ATCGTTGCAA CTGGGAGGAC AGAGACAGAA GTCAACTGTC 2340
TCCTTTGCTG CACTCCACCA CGGAGTGGGC CATTTTACCT TGCCTTACT CGGACCTGCC 2400
CGCCTTGTCG ACTGGTCTTC TCCACCTCCA CCAAAACATC GTGGACGTGC AATTCATGTA 2460
TGGCCTATCA CCTGCTCTCA CAAAATACAT CGTCCGATGG GAGTGGGTAG TACTCTTATT 2520
CCTGCTCTTA GCGGACGCCA GGGTTTGGCG CTGCTTATGG ATGCTCATCT TGTGGGCCA 2580
GGCCGAAGCA GCACTAGAGA AGTTGGTCGT CTGACCGCT GCGAGCGCAG CTAGCTGCAA 2640
TGGCTTCTA TACTTTGTCA TCTTTTTCGT GGCTGCTTGG TACATCAAGG GTCGGGTAGT 2700
CCCTTGGCT ACTTATCCC TCACTGGCCT ATGGTCCTTT GGCCTACTGC TCCTAGCATT 2760
GCCCCAACAG GCTTATGCTT ATGACGCAIC TGTACATGGT CAGATAGGAG CAGCTCTGTT 2820
GGTACTGATC ACTCTCTTCA CACTCACCCC CGGGTATAAG ACCCTTCTCA GCGGTTTCT 2880

GTGGTGGTTG TGCTATCTTC TGACCCTGGC GGAAGCTATG GTCCAGGAGT GGGCACCACC 2940
TATGCAGGTG CGCGGTGGCC GTGATGGGAT CATATGGGCC GTCGCCATAT TCTGCCCGGG 3000
TGTGGTGTTT GACATAACCA AGTGGCTCTT GCGGTGCTT GGGCCTGCTT ATCTCCTAAA 3060
AGGTGCTTTG ACGCGTGTGC CGTACTTCGT CAGGGCTCAC GCTCTACTAA GGATGTGCAC 3120
CATGGTAAGG CATCTCGCGG GGGGTAGGTA CGTCCAGATG GTGCTACTAG CCCTTGGCAG 3180
GTGGACTGGC ACTTACATCT ATGACCACCT CACCCCTATG TCGGATTGGG CTGCTAATGG 3240
CCTGCGGGAC TTGGCGGTGG CCGTGGAGCC TATCATCTTC AGTCCGATGG AGAAAAAGT 3300
CATCGTCTGG GGAGCGGAGA CAGCTGCTTG CGGGGATATC TTACACGGAC TTCCCGTGTC 3360
CGCCCGACTT GCGCGGAGG TCCTCCTTGG CCCAGCTGAT GGCTATACCT CCAAGGGGTG 3420
GAGTCTTCTC GCCCCATCA CTGCTTATGC CCAGCAGACA CGCGGCCTTT TGGGCACCAT 3480
AGTGGTAGC ATGACGGGC GCGACAAGAC AGAACAGGCC GGGGAGATTG AGGTCTGTG 3540
CACGGTCACT CAGTCTTCC TCGGAACAAC CATCTCGGGG GTCTTATGGA CTGTCTACCA 3600
TGGAGCTGGC AACAAGACTC TAGCCGGTC ACGGGTCCG GTCACACAGA TGTACTCCAG 3660
TGCTGAGGGG GACTTAGTGG GGTGGCCAG CCCCCCGGG ACCAAATCTT TGGAGCCGTG 3720
CACGTGTGGA GCGGTCGACC TATACCTGGT CAGCGGAAAC GCTGATGTCA TCCCGGCTCG 3780
AAGACCGGG GACAAGCGAG GAGCGTACT CTCCCCAGA CCTCTTTCCA CCTTGAAGGG 3840
GTCCTCGGGG GCGCCGGTGC TCTGCCCGAG AGGCCACGCT GTCGGGCTCT TCCGGGCAGC 3900
CGTGTGCTCC CGGGCGGTGG CCAAGTCCAT AGATTTTATC CCCGTTGAGA CACTTGACAT 3960
CGTCACTCGG TCCCCACCT TTAGTGACAA CAGCACACCA CCTGCTGTGC CCCAACTTA 4020
TCAGGTCGGG TACTTACATG CCCCAGCTGG TAGTGGAAAG AGCACCAAG TCCCTGTCCG 4080
GTATGCCGCT CAGGGGTACA AAGTGCTAGT GCTTAATCCC TCGGTGGCTG CCACCCTGGG 4140
GTTTGGGGCG TACTTGTCCA AGGCACATGG CATCAATCCC AACATTAGGA CTGGGGTCAG 4200
GACTGTGACG ACCGGGGCGC CCATCACGTA CTCCACATAT GGCAAATTCC TCGCCGATGG 4260
GGGCTGCGCA GCGGGCGCCT ATGACATCAT CATATGCGAT GAATGCCATG CCGTGGACTC 4320
TACCACCATT CTCGGCATCG GAACAGTCCT CGATCAAGCA GAGACAGCCG GGGTCAGGCT 4380
AACTGTACTG GCTACGGCTA CGCCCCCGG GTCAGTGACA ACCCCCCACC CCAACATAGA 4440
GGAGGTGGCC CTCGGGCAGG AGGGTGAGAT CCCCTTCTAT GGGAGGGCGA TTCCCCTGTC 4500
ATACATCAAG GGAGGAAGAC ACTTGATCTT CTGCCACTCA AAGAAAAAGT GTGACGAGCT 4560
CGCGGCGGCC CTTCGGGGTA TGGGCTTGAA CGCAGTGGCA TACTACAGAG GGCTGGACGT 4620

CTCCGTAATA CCAACTCAGG GAGACGTAGT GGTGTCGCC ACCGACGCC TCATGACGG 4680
GTTTACTGGA GACTTTGACT CCGTGATCGA CTGCAACGTA GCGGTCCTC AAGTTGTAGA 4740
CTTCAGCTTG GACCCACAT TCACCATAAC CACACAGACT GTCCCTCAAG ACGCTGTCTC 4800
ACGTAGCCAG CGCCGGGGCC GCACGGGCAG GGAAGACTG GGTATTTATA GGTATGTTT 4860
CACTGGTGAG CGAGCCTCAG GAATGTTTGA CAGTGTAGTG CTCTGCGAGT GCTACGATGC 4920
AGGGGCCGCA TGGTATGAG TCACACCAGC GGAGACCACC GTCAGGCTCA GAGCATATTT 4980
CAACACACCT GGTTCGCTG TGTGCCAAGA CCATCTGAG TTTTGGGAGG CAGTTTTCAC 5040
CGGCTCACA CACATAGATG CCCACTTCCT TTCCCAACA AAGCAATCGG GGGAAAATTT 5100
CGCATACTTA ACAGCCTACC AGGCTACAGT GTGCGCTAGG GCCAAAGCCC CCCCCCGTC 5160
CTGGGACGTC ATGTGGAAGT GTTIGACTCG ACTCAAGCCC AACTCGTGG GCCCCACACC 5220
TCTCTGTAC CCGTTGGCT CTGTTACCA CGAGGTCACC CTCACGCATC CTGTGACGAA 5280
ATACATCGCC ACCTGCATG AAGCCGACCT TGAGGTCATG ACCAGCACGT GGGTCTTAGC 5340
TGGGGGGGTC TTGGCGGCCG TCGCCGCGTA CTGCTGGCG ACCGGGTGTG TTTGCATCAT 5400
CGGCCGCTTG CACGTTAACC AGCGAGCCGT CGTTGCACCG GACAAGGAGG TCCTCTATGA 5460
GGCTTTTGAT GAGATGGAGG AATGTGCCTC TAGAGCGGCT CTCATTGAAG AGGGGCAGCG 5520
GATAGCCGAG ATGCTGAAGT CCAAGATCCA AGGCTTATTG CAGCAAGCTT CCAAACAAGC 5580
TCAAGACATA CAACCCGCTG TGCAGGCTTC TTGGCCCAAG GTAGAGCAAT TCTGGGCCAA 5640
ACACATGTGG AACTTCATCA GCGGCATTCA ATACCTCGCA GGAATATCAA CACTGCCAGG 5700
GAACCGTGT GTAGCTTCCA TGATGGCATT CAGTCCCGCC CTCACAGTC CGTTGTCAAC 5760
TAGCACCAC ATCCTTCTCA ACATTTTGGG GGGCTGGCTA GCATCCCAA TTGCGCCTCC 5820
CGCGGGGGCT ACCGCTTCG TCGTCAGTGG CTTGGTGGG GCTGCCGTAG GCAGCATAGG 5880
CTTGGGTAAG GTGCTGGTGG ACATCCTGGC AGGCTATGGT GCGGGCATT TCGGGGCTCT 5940
CGTCGCATTG AAGATCATGT CTGGCGAGAA GCCCTCCATG GAGGATGTTG TCAACCTGCT 6000
GCCTGGAATT CTGTCTCCGG GTGCCCTGGT GGTGGGAGTC ATCTGCGCG CCATCCTGCG 6060
CCGACACGTG GGACCGGGGG AAGGCCTGT CCAATGGATG AATAGGCTCA TTGCTTTGC 6120
TTCCAGAGGA AACCACGTCG CCCCACCCA CTACGTGACG GAGTCGGATG CGTCGCAGCG 6180
TGTGACCCAA CTACTGGCT CCCTTACCAT AACCAGCCTG CTCAGGAGAC TCCACAAGT 6240
GATTACTGAA GACTGCCCCA TCCATGCAG CGGCTCGTGG CTCGCGATG TGTGGGATTG 6300
GGTTTGACC ATCCTAACAG ACTTTAAAA CTGGCTGACC TCCAAATTGT TCCAAAGAT 6360

GCCTGGTCTC CCCTTTATCT CTTGTCAAAA GGGGTACAAG GCGGTGTGG CTGGCACTGG 6420
 TATCATGACC ACACGGTGTG CTTGCGGCGC CAATATCTCT GGCAATGTCC GCCTGGGCTC 6480
 CATGAGAATT ACGGGGCCCA AAACCTGCAI GAATATCTGG CAGGGGACCT TTCCCATCAA 6540
 TTGTTACAGG GAGGGCCAGT GCGTGCCGAA ACCCGCACCA AACTTTAAGA TCGCCATCTG 6600
 GAGGGTGGCG GCCTCAGAGT ACGCGGAGGT GACGCAACAC GGGTCATACC ACTACATAAC 6660
 AGGACTTACC ACTGATAACT TGAAAGTTCC TTGCCAACTA CCTTCTCCAG AGTTCTTTTC 6720
 CTGGGTGGAC GGAGTGCGA TCCATAGGTT TGCCCCATA CCGAAGCCGT TTTTTCGGGA 6780
 TGAGGTCTCG TTCTGCGTIG GGTTAATTC ATTTGTGTC GGGTCTCAGC TCCCTTGCGA 6840
 TCCTGAACCT GACACAGACG TATTGACGTC CATGCTAACA GACCCATCCC ATATCACGGC 6900
 GGAGACTGCA GCGCGGCGTT TGGCACGGGG GTACCCCCG TCCGAGGCAA GCTCCTCAGC 6960
 GAGCCAGCTA TCGGCACCAT CGCTGCGAGC CACCTGCACC ACCCAGGCA AGGCCTATGA 7020
 TGTGGACATG GTGGATGCCA ACCTGTTCAI GGGGGGCGAT GTGACCCGGA TAGAGTCTGA 7080
 GTCCAAAGTG GTCGTCTGG ACTCTCTCGA CCCAATGGTC GAAGAAAGGA GCGACCTTGA 7140
 GCCTTCGATA CCATCGGAAT ATATGCTCCC CAAGAAGAGA TTCCCACCAG CCTTACCGGC 7200
 TTGGGCACGG CCTGATTACA ACCCACCCT TGTGGAATCG TGAAGAGGC CAGATTACCA 7260
 ACCGGCCACT GTTGCGGCT GCGCTCTCCC CCCCCTAAG AAAACCCGA CGCCTCCCC 7320
 AAGGAGACGC CGGACAGTGG GTCTGAGTGA GAGTCCATA GCAGATGCCC TACAACAGCT 7380
 GGCATCAAG TCCTTTGGCC AGCCCCCCCC AAGCGCGAT TCAGGCCTTT CCACGGGGGC 7440
 GGAGCGAGCC GATTCCGGCA GTCCGACGCC CCCCAGTGG TTGGCCCTTT CGGAGACAGG 7500
 TTCCATCTCC TCCATGCCCC CTCTCGAGGG GGAGCCTGGA GATCCAGACT TGGAGCCTGA 7560
 GCAGGTAGAG CTTCACCTC CCCCCAGGG GGGGGTGGTA ACCCCCGCT CAGGCTCGGG 7620
 GTCTTGGTCT ACTTGCTCCG AGGAGGACGA CTCGTCGTG TGCTGCTCCA TGTCATACTC 7680
 CTGGACCGGG GCTCTAATAA CTCCTTGTAG CCCCAGAGAG GAAAAGTTGC CAATTGGCCC 7740
 CTTGAGCAAC TCCCTGTTGC GATATCACA CAAGGTGTAC TGTACCACAT CAAAGAGCGC 7800
 CTCATTAAGG GCTAAAAAGG TAACITTTGA TAGGATGCAA GCGCTCGACG CTCATTATGA 7860
 CTCAGTCTTG AAGGACATTA AGCTAGCGGC CTCCAAGGTC ACCGCAAGGC TTCTCACITT 7920
 AGAGGAGGCC TGCCAGTTAA CTCACCCCCA CTCTGCAAGA TCCAAGTATG GGTITGGGGC 7980
 TAAGGAGGTC CGCAGCTTGT CCGGGAGAGC CGTTAACCAC ATCAAGTCCG TGTGGAAGGA 8040
 CCTCCTGGAA GACACACAAA CACCAATTCC TACAACCATC ATGCCAAAA ATGAGGTGTT 8100

CTGCGTGGAC CCCACCAAGG GGGGTAAGAA AGCAGCTCGC CTTATCGTTT ACCCTGACCT 8160
 CGGCGTCAGG GTCTGCGAGA AAATGGCCCT TTATGATATC ACACAAAAGC TTCCTCAGGC 8220
 GGTGATGGGG GCTTCTTATG GATTCCAGTA CTCCCCCGCT CAGCGGGTGG AGTTTCTCTT 8280
 GAAGGCATGG GCGGAAAAGA AAGACCCTAT GGGTTTTTCG TATGATACCC GATGCTTTGA 8340
 CTCAACCGTC ACTGAGAGAG ACATCAGGAC TGAGGAGTCC ATATATCGGG CTTGTTCCCT 8400
 GCGCGAGGAG GCCCAGACTG CCATACACTC ACTGACTGAG AGACTTTACG TGGGAGGGCC 8460
 CATGTTCAAC AGCAAGGGCC AGACCTGCGG GTACAGGCGT TGCCGCGCCA GCGGGGTGCT 8520
 TACCACTAGC ATGGGGAACA CCATCACATG CTATGTGAAA GCCTTAGCGG CCTGTAAGGC 8580
 TGCAGGGATA ATTGCGCCCA CAATGCTGGT ATGCGGGGAT GACTTGGTGG TCATCTCAGA 8640
 GAGCCAGGGG ACCGAGGAGG ACGAGCGGAA CCTGAGAGCC TTCACGGAGG CTATGACCAG 8700
 GTATTCTGCC CCTCCTGGTG ACCCCCCAG ACCGGAATAT GACCTGGAGC TGATAACATC 8760
 TTGCTCCTCA AATGTGTCTG TGGCGTTGGG CCCACAAGGC CGCCGCAGAT ACTACCTGAC 8820
 CAGAGACCCT ACCACTCCAA TCGCCCGGGC TGCCTGGGAA ACAGTTAGAC ACTCCCTGT 8880
 CAATTCATGG CTAGGAAACA TCATCCAGTA CGCCCCAACC ATATGGGCTC GCATGGTCCT 8940
 GATGACACAC TTCTTCTCCA TTCTCATGGC CCAAGATACT CTGGACCAGA ACCTCAACTT 9000
 TGAGATGTAC GGAGCGGTGT ACTCCGTGAG TCCCTTGGAC CTCCAGCCA TAATTGAAAG 9060
 GTTACACGGG CTTGACGCTT TCTCTCTGCA CACATACACT CCCACGAAC TGACACGGGT 9120
 GGCTTCAGCC CTCAGAAAAC TTGGGGCGCC ACCCCTCAGA GCGTGAAGA GCCGGGCACG 9180
 TGCAGTCAGG GCGTCCCTCA TCTCCGTGG GGGGAGAGCG GCCGTTGCG GCCGATATCT 9240
 CTTCAACTGG GCGGTGAAGA CCAAGCTCAA ACTCACTCCA TTGCCGGAAG CGCGCTCCT 9300
 GGATTIATCC AGCTGGTTCA CTGTCGGCGC CGCGGGGGGC GACATTTATC ACAGCGTGTC 9360
 GCGTGCCCGA CCGCGCTTAT TACTCCTTGG CCTACTCCTA CTTTTGTAG GGGTAGGCCT 9420
 TTTCTACTC CCGCTCGGT AGAGCGGCAC ACATTAGCTA CACTCCATAG CTAAGTGTCC 9480
 CTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 9540
 TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 9589

【0023】配列番号：3

配列の長さ：3970

配列の型：核酸

鎖の数：一本鎖

トポロジー：直鎖状

配列の種類：cDNA to genomic RNA

特徴を決定した方法：E

40 配列

GGCATTACCC CTGCACAGTT AACTATACCA TCTTCAAAT AAGGATGTAT GTGGGGGGGG 60
TCGAGCACAG GCTCACGGCT GCGTGCAATT TCACTCGTGG GGATCGTTGC AACTTGGAGG 120
ACAGAGACAG AAGTCAACTG TCTCCTTTGC TGCCTCCAC CACGGAGTGG GCCATTTTAC 180
CTTGCACTTA CTCGGACCTG CCCGCCCTGT CGACTGGTCT TCTCCACCTC CACCAAAACA 240
TCGTGGACGT GCAATTCATG TATGGCCTAT CACCTGCTCT CACAAAATAC ATCGTCCGAT 300
GGGAGTGGGT AGTACTCTTA TTCCTGCTCT TAGCGGACGC CAGGGTTTGC GCCTGCTTAT 360
GGATGCTCAT CTGTGTGGGC CAGGCCGAAG CAGCACTAGA GAAGTTGGTC GTCTTGACAG 420
CTGCGAGCGC AGCTAGCTGC AATGGCTTCC TATACTTTGT CATCTTTTTC GTGGCTGCTT 480
GGTACATCAA GGGTCGGGTA GTCCCCCTGG CTACTTATTC CCTCACTGGC CTATGGTCTT 540
TTGGCCTACT GCTCCTAGCA TTGCCCAAC AGGCTTATGC TTATGACGCA TCTGTACATG 600
GTCAGATAGG AGCAGCTCTG TTGGTACTGA TCACTCTCTT TACACTCACC CCCGGGTATA 660
AGACCCTTCT CAGCCGGTTT CTGTGGTGGT TGTGCTATCT TCTGACCCTG GCGGAAGCTA 720
TGGTCCAGGA GTGGGCACCA CCTATGCAGG TGCGCCGTGG CCGTGATGGG ATCATATGGG 780
CCGTGCCCAT ATTCTGCCCG GGTGTGGTGT TTGACATAAC CAAGTGGCTC TTGGCGGTGC 840
TTGGGCCTGC TTATCTCCTA AAAGGTGCTT TGACGCGTGT GCCGTACTTC GTCAGGGCTC 900
ACGCTCTACT AAGGATGTGC ACCATGGTAA GGCATCTCGC GGGGGGTAGG TACGTCCAGA 960
TGGTGCTACT AGCCCTTGGC AAGTGGACTG GCCTTACAT CTATGACCAC CTCACCCCTA 1020
TGTCGGATTG GGCTGCTAAT GGCCTGCGGG ACTTGGCGGT CGCCGTGGAG CCTATCATCT 1080
TCAGTCCGAT GGAGAAAAA GTCATCGTCT GGGGAGCGGA GACAGCTGCT TCGGGGGATA 1140

TCTTACACGG ACTTCCCGTG TCCGCCCGAC TTGGCCGGGA GGTCTCTCTT GGGCCAGCTG 1200
ATGGCTATAC CTCCAAGGGG TGGAGTCTTC TCGCCCCCAT CACTGCTTAT GGGCAGCAGA 1260
CACGCGGCCT TTTGGGCACC ATAGTGGTGA GCATGACGGG GCGCGACAAG ACAGAACAGG 1320
CCGGGGAGAT TCAGGTCCTG TCCACGGTCA CTCAGTCCCT CCTCGGAACA ACCATCTCGG 1380
GGGTCTTATG GACTGTCTAC CATGGAGCTG GCAACAAGAC TCTAGCCGGC TCACGGGGTC 1440
CGGTACACAA GATGTACTCC AGTGTGAGG GGGACTTAGT GGGGTGGCCC AGCCCCCCCC 1500
GGACCAAATC TTTGGAGCCG TGCACGTGTG GAGCGGTGGA CCTATACCTG GTCACGCGAA 1560
ACGCTGATGT CATCCCGGCT CGAAGACGCG GGGACAAGCG AGGAGCGCTA CTCTCCCCGA 1620
GACCTCTTTC CACCTTGAAG GGGTCTCGG GGGGCCCGGT GCTCTGCCCC AGAGGCCACG 1680
CTGTGCGGGT CTTCGGGCA GCCGTGTGCT CCCGGGGCGT GGGCAAGTCC ATAGATTTTA 1740
TCCCGTTGA GACACTTGAC ATCGTCACTC GGTCCCCAC CTTAGTGAC AACAGCACAC 1800
CACCTGCTGT GGGCCAACT TATCAGGTG GGTACTTACA TGCCCCGACT GGTAGTGGA 1860
AGAGCACCAA AGTCCCTGTC GGTATGCGC CTCAGGGGA CAAAGTGCTA GTGCTTAATC 1920
CCTCGGTGGC TGCCACCCTG GGGTTTGGG CGTACTTGT CAAAGGACAT GGCATCAATC 1980
CCAACATTAG GACTGGGGTC AGGACTGTGA CGACCGGGC GCCCATCAG TACTCCACAT 2040
ATGGCAAATT CCTCGCGAT GGGGGCTGCG CAGGCGGGC CTATGACATC ATCATATGCG 2100
ATGAATGCCA TGCCGTGGAC TCTACCACCA TTCTCGGCAT CGGAACAGTC CTCGATCAAG 2160
CAGAGACAGC CGGGGTGAG CTAAGTGTAC TGGCTACGGC TACGCCCCC GGGTCAGTGA 2220
CAACCCCCCA CCGCAACATA GAGGAGGTGG CCCTCGGCA GGAGGGTGAG ATCCCTTCT 2280
ATGGGAGGGC GATTCCCTG TCATACATCA AGGAGGAAG ACACTTGATC TTCTGCCACT 2340
CAAAGAAAAA GTGTGACGAG CTCGCGGGCG CCCTCGGGG TATGGGCTTG AACCGAGTGG 2400
CATACTACAG AGGGCTGGAC GTCTCGTAA TACCAACTCA GGGAGACGTA GTGGTCGTG 2460
CCACCGACGC CCTCATGAG GGGTTTACTG GAGACTTTGA CTCGGTATC GACTGCAACG 2520
TAGGGGTGAC TCAAGTTGTA GACTTCAGCT TGGACCCAC ATTACCCATA ACCACACAGA 2580
CTGTCCCTCA AGACGCTGTC TCACGTAGCC AGCGCCGGG CCGCACGGG AGGGGAAGAC 2640
TGGGTATTTA TAGGTATGTT TCCACTGGTG AGCGAGCCTC AGGAATGTTT GACAGTGTAG 2700
TGCTCTGCGA GTGCTACGAT GCAGGGGGCG CATGGTATGA GCTCACACCA GCGGAGACCA 2760
CCGTGAGGT CAGAGCATAT TTCAACACAC CTGGTTTCC TGTGTGCCAA GACCATCTTG 2820
AGTTTTGGGA GCAGTTTTC ACCGGCTCA CACACATAGA TGCCCACTC CTTCCTCCAA 2880

CAAAGCAATC GGGGGAAAAT TTCGCATACT TAACAGCCTA CCAGGCTACA GTGTGCGCTA 2940
 GGGCCAAAGC CCCCCCCCCG TCCTGGGACG TCATGTGGAA GTGTTTGACT CGACTCAAGC 3000
 CCACACTCGT GGGCCCCACA CCTCTCCTGT ACCGCTTGGG CTCTGTTACC AACGAGGTCA 3060
 CCCTCACGCA TCCTGTGACG AAATACATCG CCACCTGCAT GCAAGCCGAC CTTGAGGTCA 3120
 TGACCAGCAC GTGGGTCTTA GCTGGGGGGG TCTTGGCGGC CGTCGCCGCG TACTGCCTGG 3180
 CGACCGGGTG TGTTTGCATC ATCGGCCGCT TGCACGTTAA CCAGCGAGCC GTCGTTGCAC 3240
 CGGACAAGGA GGTCTCTAT GAGGCTTTTG ATGAGATGGA GGAATGTGCC TCTAGAGCGG 3300
 CTCTCATTGA AGAGGGGCGAG CGGATAGCCG AGATGCTGAA GTCCAAGATC CAAGGCTTAT 3360
 TGCAGCAAGC TTCCAACAA GCTCAAGACA TACAACCCGC TGTGCAGGCT TCTTGGCCCA 3420
 AGGTAGAGCA ATTCTGGGCC AAACACATGT GGAACCTCAT CAGCGGCATT CAATACCTCG 3480
 CAGGACTATC AACACTGCCA GGAACCCCTG CTGTAGCTTC CATGATGGCA TTCAGTGCCG 3540
 CCCTCACCAG TCGGTTGTCA ACTAGCACCA CTATCCTTCT CAACATTTTG GGGGGCTGGC 3600
 TAGCATCCCA AATTGCGCCT CCCGCGGGGG CTACCGGCTT CGTCGTCACT GGCCTGGTGG 3660
 GGGCTGCCGT AGGCAGCATA GGCTTGGGTA AGGTGCTGGT GGACATCCTG GCAGGGTATG 3720
 GTGCGGGCAT TTCGGGGGCT CTCGTGGCAT TCAAGATCAT GTCTGGCGAG AAGCCCTCCA 3780
 TGGAGGATGT TGTCAACCTG CTGCCCTGAA TTCTGTCTCC GGGTGGCCTG GTGGTGGGAG 3840
 TCATCTGCGC GGCCATCCTG CGCCGACACG TGGGACCGGG GGAAGGCGCT GTCCAATGGA 3900
 TGAATAGGCT CATTGCCTTT GCTTCCAGAG GAAACCACGT CGCCCCCACC CACTACGTGA 3960
 CGGAGTCGGA 3970

【0024】配列番号：4

配列の長さ：2693

配列の型：核酸

鎖の数：一本鎖

トポロジー：直鎖状

配列の種類：cDNA to genomic RNA

30 特徴を決定した方法：E

配列

ATTCTGTCTC CGGGTGGCCT GGTGGTGGGA GTCATCTGCG CGGCCATCCT GCGCCGACAC 60
GTGGGACCGG GGAAGGCGC TGTCCAATGG ATGAATAGGC TCATTGCCYT TGCTTCCAGA 120
GGAAACCACG TCGCCCCAC CCACTACGTG ACGGAGTCGG ATGCGTCGCA GCGTGTGACC 180
CAACTACTTG GCTCCCTTAC CATAACCAGC CTGCTCAGGA GACTCCACAA CTGGATTACT 240
GAAGACTGCC CCATCCCATG CAGCGGCTCG TGGCTCCGCG ATGTGTGGGA TTGGGTTTGC 300
ACCATCCTAA CAGACTTTAA AACTGGCTG ACCTCCAAAT TGTTCCCAA GATGCCCTGT 360
CTCCCCTTA TCTCTTGTC AAGGGGTAC AAGGGCGTGT GGGCTGGCAC TGGTATCATG 420
ACCACACGGT GTCCTTGCGG CGCCAATATC TCTGGCAATG TCCGCTGGG CTCCATGAGA 480
ATTACGGGGC CAAAACCTG CATGAATATC TGGCAGGGGA CCTTCCCAT CAATTGTTAC 540
ACGGAGGGCC AGTGCGTGC GAAACCCGCA CAAACTTTA AGATGCCAT CTGGAGGCTG 600
GCGGCTCAG AGTACCGGA GGTGACGCG CACGGGTCAT ACCACTACAT AACAGGACTT 660
ACCACTGATA ACTTGAAAGT TCCTTGCCAA CTACCTTCTC CAGAGTTCTT TTCCTGGGTG 720
GACGGAGTGC AGATCCATAG GTTTGCCCC ATACCGAAGC CGTTTTTTCG GGATGAGGTC 780
TCGTTCTGCG TTGGGCTTAA TTCATTTGTC GTCGGGTCTC AGCTCCCTTG CGATCCTGAA 840
CCTGACACAG ACGTATTGAC GTCCATGCTA ACAGACCCAT CCCATATCAC GCGGGAGACT 900
GCAGCGCGGC GTTTGGCAG GGGGTCACCC CCGTCCGAG CAAGCTCCTC AGCGAGCCAG 960
CTATCGGCAC CATCGCTGCG AGCCACCTGC ACCACCCACG GCAAGGCCTA TGATGTGGAC 1020
ATGGTGGATG CCAACCTGTT CATGGGGGGC GATGTGACCC GGATAGAGTC TGAGTCCAAA 1080
GTGTCGTTT TGGACTCTCT CGACCCAATG GTCGAAGAAA GGAGCGACCT TGAGCCTTCG 1140

ATACCATCGG AATATATGCT CCCCAAGAAG AGATTCCCAC CAGCCTTACC GGCTTGGGCA 1200
 CGGCCTGATT ACAACCCACC GCTTGTGGAA TCGTGGAGA GGCCAGATTA CCAACCGGCC 1260
 ACTGTTGCGG GCTGCGCTCT CCCCCCCCCT AAGAAAACCC CGACGCCTCC CCCAAGGAGA 1320
 CGCCGGACAG TGGGTCTGAG TGAGAGCTCC ATAGCAGATG CCCTACAACA GCTGGCCATC 1380
 AAGTCCTTTG GCCAGCCCCC CCCAAGCGGC GATTACGGCC TTTCACGGG GCGGGACGCA 1440
 GCCGATTCCG GCAGTCGGAC GCCCCCGAT GAGTGGGCC TTTCGGAGAC AGGTTCCATC 1500
 TCCTGCATGC CCCCTCTCGA GGGGGAGCCT GGAGATCCAG ACTTGGAGCC TGAGCAGGTA 1560
 GAGCTTCAAC CTCCCCCCA GGGGGGGGTG GTAACCCCGG GCTCAGGCTC GGGGTCTTGG 1620
 TCTACTTGCT CCGAGGAGGA CGACTCCGTC GTGTGCTGCT CCATGTCATA CTCCTGGACC 1680
 GGGGCTCTAA TAACTCCTTG TAGCCCCGAA GAGGAAAAGT TGCCAATTGG CCCCTTGAGC 1740
 AACTCCCTGT TGGGATATCA CAACAAGGTG TACTGTACCA CATCAAAGAG CGCCTCATT 1800
 AGGGCTAAAA AGGTAACCTT TGATAGGATG CAAGCGCTCG ACGCTCATT 1860
 TTGAAGGACA TTAAGCTAGC GGCCTCCAAG GTCACCGCAA GGCTTCTCAC TTTAGAGGAG 1920
 GCCTGCCAGT TAACTCCACC CCACTCTGCA AGATCCAAGT ATGGGTTTGG GGCTAAGGAG 1980
 GTCCGCAGCT TGTCCGGGAG AGCCGTTAAC CACATCAAGT CCGTGTGGAA GGACCTCCTG 2040
 GAAGACACAC AAACACCAAT TCCTACAACC ATCATGGCCA AAAATGAGGT GTTCTGCGTG 2100
 GACCCACCA AGGGGGGTAA GAAAGCAGCT CGCCTTATCG TTTACCCTGA CCTCGGGGTC 2160
 AGGGTCTGCG AGAAAATGGC CCTTTATGAT ATCACACAAA AGCTTCCTCA GCGGGTGATG 2220
 GGGGCTTCTT ATGGATTCCA GACTCCCCC GCTCAGCGG TGGAGTTTCT CTTGAAGGCA 2280
 TGGGCGGAAA AGAAAGACCC TATGGGTTT TCGTATGATA CCGATGCTT TGAICTAACC 2340
 GTCAGTAGA GAGACATCAG GACTGAGGAG TCCATATATC GGGCTTGTTC CTTGCCCGAG 2400
 GAGGCCACAC CTGCCATACA CTCAGTACT GAGAGACTTT ACGTGGGAGG GCCCATGTTT 2460
 AACAGCAAGG GCCAGACCTG CGGGTACAGG CGTTGCCGCG CCAGCGGGGT GCTTACCACT 2520
 AGCATGGGGA ACACCATCAC ATGCTATGTG AAAGCCTTAG CGGCCTGTAA GGCTGCAGGG 2580
 ATAATTGCGC CCACAATGCT GGTATGCGGC GATGACTTGG TTGTCATCTC AGAGAGCCAG 2640
 GGGACCGAGG AGGACGAGCG GAACCTGAGA GCCTTCACGG AGGCTATGAC CAG 2693

【0025】配列番号：5

配列の長さ：3033

配列の型：アミノ酸

トポロジー：直鎖状

40 配列の種類：蛋白質

配列

Met	Ser	Thr	Asn	Pro	Lys	Pro	Gln	Arg	Lys	Thr	Lys	Arg	Asn	Thr
				5					10					15
Asn	Arg	Arg	Pro	Gln	Asp	Val	Lys	Phe	Pro	Gly	Gly	Gly	Gln	Ile
				20					25					30
Val	Gly	Gly	Val	Tyr	Leu	Leu	Pro	Arg	Arg	Gly	Pro	Arg	Leu	Gly
				35					40					45
Val	Arg	Ala	Thr	Arg	Lys	Thr	Ser	Glu	Arg	Ser	Gln	Pro	Arg	Gly
				50					55					60
Arg	Arg	Gln	Pro	Ile	Pro	Lys	Asp	Arg	Arg	Ser	Thr	Gly	Lys	Ser
				65					70					75
Trp	Gly	Lys	Pro	Gly	Tyr	Pro	Trp	Pro	Leu	Tyr	Gly	Asn	Glu	Gly
				80					85					90
Leu	Gly	Trp	Ala	Gly	Trp	Leu	Leu	Ser	Pro	Arg	Gly	Ser	Arg	Pro
				95					100					105
Ser	Trp	Gly	Pro	Asn	Asp	Pro	Arg	His	Arg	Ser	Arg	Asn	Val	Gly
				110					115					120
Lys	Val	Ile	Asp	Thr	Leu	Thr	Cys	Gly	Phe	Ala	Asp	Leu	Met	Gly
				125					130					135
Tyr	Ile	Pro	Val	Val	Gly	Ala	Pro	Leu	Gly	Gly	Val	Ala	Arg	Ala
				140					145					150
Leu	Ala	His	Gly	Val	Arg	Val	Leu	Glu	Asp	Gly	Val	Asn	Phe	Ala

155	160	165
Thr Gly Asn Leu Pro Gly Cys Ser Phe	Ser Ile Phe Leu Leu Ala	
170	175	180
Leu Leu Ser Cys Ile Thr Thr Pro Val	Ser Ala Ala Glu Val Lys	
185	190	195
Asn Ile Ser Thr Gly Tyr Met Val Thr	Asn Asp Cys Thr Asn Asp	
200	205	210
Ser Ile Thr Trp Gln Leu Gln Ala Ala Val	Leu His Val Pro Gly	
215	220	225
Cys Val Pro Cys Glu Lys Val Gly Asn Thr	Ser Arg Cys Trp Ile	
230	235	240
Pro Val Ser Pro Asn Val Ala Val Gln Gln	Pro Gly Ala Leu Thr	
245	250	255
Gln Gly Leu Arg Thr His Ile Asp Met Val	Val Met Ser Ala Thr	
260	265	270
Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu	Cys Gly Gly Val Met	
275	280	285
Leu Ala Ala Gln Met Phe Ile Val Ser Pro	Gln His His Trp Phe	
290	295	300
Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro	Gly Thr Ile Thr Gly	
305	310	315
His Arg Met Ala Trp Asp Met Met Met	Asn Trp Ser Pro Thr Ala	
320	325	330
Thr Met Ile Leu Ala Tyr Ala Met Arg Val	Pro Glu Val Ile Ile	
335	340	345
Asp Ile Ile Gly Gly Ala His Trp Gly Val	Met Phe Gly Leu Ala	
350	355	360
Tyr Phe Ser Met Gln Gly Ala Trp Ala Lys	Val Val Val Ile Leu	
365	370	375

Leu Leu Ala Ala Gly Val Asp Ala Gln Thr His Thr Val Gly Gly	380	385	390
Ser Thr Ala His Asn Ala Arg Thr Leu Thr Gly Met Phe Ser Leu	395	400	405
Gly Ala Arg Gln Lys Ile Gln Leu Ile Asn Thr Asn Gly Ser Trp	410	415	420
His Ile Asn Arg Thr Ala Leu Asn Cys Asn Asp Ser Leu His Thr	425	430	435
Gly Phe Leu Ala Ser Leu Phe Tyr Thr His Ser Phe Asn Ser Ser	440	445	450
Gly Cys Pro Glu Arg Met Ser Ala Cys Arg Ser Ile Glu Ala Phe	455	460	465
Arg Val Gly Trp Gly Ala Leu Gln Tyr Glu Asp Asn Val Thr Asn	470	475	480
Pro Glu Asp Met Arg Pro Tyr Cys Trp His Tyr Pro Pro Arg Gln	485	490	495
Cys Gly Val Val Ser Ala Ser Ser Val Cys Gly Pro Val Tyr Cys	500	505	510
Phe Thr Pro Ser Pro Val Val Val Gly Thr Thr Asp Arg Leu Gly	515	520	525
Ala Pro Thr Tyr Thr Trp Gly Glu Asn Glu Thr Asp Val Phe Leu	530	535	540
Leu Asn Ser Thr Arg Pro Pro Gln Gly Ser Trp Phe Gly Cys Thr	545	550	555
Trp Met Asn Ser Thr Gly Tyr Thr Lys Thr Cys Gly Ala Pro Pro	560	565	570
Cys Arg Ile Arg Ala Asp Phe Asn Ala Ser Met Asp Leu Leu Cys	575	580	585
Pro Thr Asp Cys Phe Arg Lys His Pro Asp Thr Thr Tyr Ile Lys			

590	595	600
Cys Gly Ser Gly Pro Trp Leu Thr Pro Arg Cys Leu Ile Asp Tyr		
605	610	615
Pro Tyr Arg Leu Trp His Tyr Pro Cys Thr Val Asn Tyr Thr Ile		
620	625	630
Phe Lys Ile Arg Met Tyr Val Gly Gly Val Glu His Arg Leu Thr		
635	640	645
Ala Ala Cys Asn Phe Thr Arg Gly Asp Arg Cys Asn Leu Glu Asp		
650	655	660
Arg Asp Arg Ser Gln Leu Ser Pro Leu Leu His Ser Thr Thr Glu		
665	670	675
Trp Ala Ile Leu Pro Cys Thr Tyr Ser Asp Leu Pro Ala Leu Ser		
680	685	690
Thr Gly Leu Leu His Leu His Gln Asn Ile Val Asp Val Gln Phe		
695	700	705
Met Tyr Gly Leu Ser Pro Ala Leu Thr Lys Tyr Ile Val Arg Trp		
710	715	720
Glu Trp Val Val Leu Leu Phe Leu Leu Leu Ala Asp Ala Arg Val		
725	730	735
Cys Ala Cys Leu Trp Met Leu Ile Leu Leu Gly Gln Ala Glu Ala		
740	745	750
Ala Leu Glu Lys Leu Val Val Leu His Ala Ala Ser Ala Ala Ser		
755	760	765
Cys Asn Gly Phe Leu Tyr Phe Val Ile Phe Phe Val Ala Ala Trp		
770	775	780
Tyr Ile Lys Gly Arg Val Val Pro Leu Ala Thr Tyr Ser Leu Thr		
785	790	795
Gly Leu Trp Ser Phe Gly Leu Leu Leu Leu Ala Leu Pro Gln Gln		
800	805	810

Ala Tyr Ala Tyr Asp Ala Ser Val His Gly Gln Ile Gly Ala Ala			
	815	820	825
Leu Leu Val Leu Ile Thr Leu Phe Thr Leu Thr Pro Gly Tyr Lys			
	830	835	840
Thr Leu Leu Ser Arg Phe Leu Trp Trp Leu Cys Tyr Leu Leu Thr			
	845	850	855
Leu Ala Glu Ala Met Val Gln Glu Trp Ala Pro Pro Met Gln Val			
	860	865	870
Arg Gly Gly Arg Asp Gly Ile Ile Trp Ala Val Ala Ile Phe Cys			
	875	880	885
Pro Gly Val Val Phe Asp Ile Thr Lys Trp Leu Leu Ala Val Leu			
	890	895	900
Gly Pro Ala Tyr Leu Leu Lys Gly Ala Leu Thr Arg Val Pro Tyr			
	905	910	915
Phe Val Arg Ala His Ala Leu Leu Arg Met Cys Thr Met Val Arg			
	920	925	930
His Leu Ala Gly Gly Arg Tyr Val Gln Met Val Leu Leu Ala Leu			
	935	940	945
Gly Arg Trp Thr Gly Thr Tyr Ile Tyr Asp His Leu Thr Pro Met			
	950	955	960
Ser Asp Trp Ala Ala Asn Gly Leu Arg Asp Leu Ala Val Ala Val			
	965	970	975
Glu Pro Ile Ile Phe Ser Pro Met Glu Lys Lys Val Ile Val Trp			
	980	985	990
Gly Ala Glu Thr Ala Ala Cys Gly Asp Ile Leu His Gly Leu Pro			
	995	1000	1005
Val Ser Ala Arg Leu Gly Arg Glu Val Leu Leu Gly Pro Ala Asp			
	1010	1015	1020
Gly Tyr Thr Ser Lys Gly Trp Ser Leu Leu Ala Pro Ile Thr Ala			

1025	1030	1035
Tyr Ala Gln Gln Thr Arg Gly Leu Leu Gly Thr Ile Val Val Ser		
1040	1045	1050
Met Thr Gly Arg Asp Lys Thr Glu Gln Ala Gly Glu Ile Glu Val		
1055	1060	1065
Leu Ser Thr Val Thr Gln Ser Phe Leu Gly Thr Thr Ile Ser Gly		
1070	1075	1080
Val Leu Trp Thr Val Tyr His Gly Ala Gly Asn Lys Thr Leu Ala		
1085	1090	1095
Gly Ser Arg Gly Pro Val Thr Gln Met Tyr Ser Ser Ala Glu Gly		
1100	1105	1110
Asp Leu Val Gly Trp Pro Ser Pro Pro Gly Thr Lys Ser Leu Glu		
1115	1120	1125
Pro Cys Thr Cys Gly Ala Val Asp Leu Tyr Leu Val Thr Arg Asn		
1130	1135	1140
Ala Asp Val Ile Pro Ala Arg Arg Arg Gly Asp Lys Arg Gly Ala		
1145	1150	1155
Leu Leu Ser Pro Arg Pro Leu Ser Thr Leu Lys Gly Ser Ser Gly		
1160	1165	1170
Gly Pro Val Leu Cys Pro Arg Gly His Ala Val Gly Val Phe Arg		
1175	1180	1185
Ala Ala Val Cys Ser Arg Gly Val Ala Lys Ser Ile Asp Phe Ile		
1190	1195	1200
Pro Val Glu Thr Leu Asp Ile Val Thr Arg Ser Pro Thr Phe Ser		
1205	1210	1215
Asp Asn Ser Thr Pro Pro Ala Val Pro Gln Thr Tyr Gln Val Gln		
1220	1225	1230
Tyr Leu His Ala Pro Thr Gly Ser Gly Lys Ser Thr Lys Val Pro		
1235	1240	1245

Val Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro		
1250	1255	1260
Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Leu Ser Lys Ala		
1265	1270	1275
His Gly Ile Asn Pro Asn Ile Arg Thr Gly Val Arg Thr Val Thr		
1280	1285	1290
Thr Gly Ala Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala		
1295	1300	1305
Asp Gly Gly Cys Ala Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp		
1310	1315	1320
Glu Cys His Ala Val Asp Ser Thr Thr Ile Leu Gly Ile Gly Thr		
1325	1330	1335
Val Leu Asp Gln Ala Glu Thr Ala Gly Val Arg Leu Thr Val Leu		
1340	1345	1350
Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Thr Pro His Pro Asn		
1355	1360	1365
Ile Glu Glu Val Ala Leu Gly Gln Glu Gly Glu Ile Pro Phe Tyr		
1370	1375	1380
Gly Arg Ala Ile Pro Leu Ser Tyr Ile Lys Gly Gly Arg His Leu		
1385	1390	1395
Ile Phe Cys His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Ala		
1400	1405	1410
Leu Arg Gly Met Gly Leu Asn Ala Val Ala Tyr Tyr Arg Gly Leu		
1415	1420	1425
Asp Val Ser Val Ile Pro Thr Gln Gly Asp Val Val Val Val Ala		
1430	1435	1440
Thr Asp Ala Leu Met Thr Gly Phe Thr Gly Asp Phe Asp Ser Val		
1445	1450	1455
Ile Asp Cys Asn Val Ala Val Thr Gln Val Val Asp Phe Ser Leu		

1460	1465	1470
Asp Pro Thr Phe Thr Ile Thr Thr Gln Thr Val Pro Gln Asp Ala		
1475	1480	1485
Val Ser Arg Ser Gln Arg Arg Gly Arg Thr Gly Arg Gly Arg Leu		
1490	1495	1500
Gly Ile Tyr Arg Tyr Val Ser Thr Gly Glu Arg Ala Ser Gly Met		
1505	1510	1515
Phe Asp Ser Val Val Leu Cys Glu Cys Tyr Asp Ala Gly Ala Ala		
1520	1525	1530
Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg Ala		
1535	1540	1545
Tyr Phe Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu		
1550	1555	1560
Phe Trp Glu Ala Val Phe Thr Gly Leu Thr His Ile Asp Ala His		
1565	1570	1575
Phe Leu Ser Gln Thr Lys Gln Ser Gly Glu Asn Phe Ala Tyr Leu		
1580	1585	1590
Thr Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Lys Ala Pro Pro		
1595	1600	1605
Pro Ser Trp Asp Val Met Trp Lys Cys Leu Thr Arg Leu Lys Pro		
1610	1615	1620
Trp Leu Val Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ser Val		
1625	1630	1635
Thr Asn Glu Val Thr Leu Thr His Pro Val Thr Lys Tyr Ile Ala		
1640	1645	1650
Thr Cys Met Gln Ala Asp Leu Glu Val Met Thr Ser Thr Trp Val		
1655	1660	1665
Leu Ala Gly Gly Val Leu Ala Ala Val Ala Ala Tyr Cys Leu Ala		
1670	1675	1680

Thr Gly Cys Val Cys Ile Ile Gly Arg Leu His Val Asn Gln Arg	1685	1690	1695
Ala Val Val Ala Pro Asp Lys Glu Val Leu Tyr Glu Ala Phe Asp	1700	1705	1710
Glu Met Glu Glu Cys Ala Ser Arg Ala Ala Leu Ile Glu Glu Gly	1715	1720	1725
Gln Arg Ile Ala Glu Met Leu Lys Ser Lys Ile Gln Gly Leu Leu	1730	1735	1740
Gln Gln Ala Ser Lys Gln Ala Gln Asp Ile Gln Pro Ala Val Gln	1745	1750	1755
Ala Ser Trp Pro Lys Val Glu Gln Phe Trp Ala Lys His Met Trp	1760	1765	1770
Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu	1775	1780	1785
Pro Gly Asn Pro Ala Val Ala Ser Met Met Ala Phe Ser Ala Ala	1790	1795	1800
Leu Thr Ser Pro Leu Ser Thr Ser Thr Thr Ile Leu Leu Asn Ile	1805	1810	1815
Leu Gly Gly Trp Leu Ala Ser Gln Ile Ala Pro Pro Ala Gly Ala	1820	1825	1830
Thr Gly Phe Val Val Ser Gly Leu Val Gly Ala Ala Val Gly Ser	1835	1840	1845
Ile Gly Leu Gly Lys Val Leu Val Asp Ile Leu Ala Gly Tyr Gly	1850	1855	1860
Ala Gly Ile Ser Gly Ala Leu Val Ala Phe Lys Ile Met Ser Gly	1865	1870	1875
Glu Lys Pro Ser Met Glu Asp Val Val Asn Leu Leu Pro Gly Ile	1880	1885	1890
Leu Ser Pro Gly Ala Leu Val Val Gly Val Ile Cys Ala Ala Ile			

1895	1900	1905
Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala Val Gln Trp Met		
1910	1915	1920
Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His Val Ala Pro		
1925	1930	1935
Thr His Tyr Val Thr Glu Ser Asp Ala Ser Gln Arg Val Thr Gln		
1940	1945	1950
Leu Leu Gly Ser Leu Thr Ile Thr Ser Leu Leu Arg Arg Leu His		
1955	1960	1965
Asn Trp Ile Thr Glu Asp Cys Pro Ile Pro Cys Ser Gly Ser Trp		
1970	1975	1980
Leu Arg Asp Val Trp Asp Trp Val Cys Thr Ile Leu Thr Asp Phe		
1985	1990	1995
Lys Asn Trp Leu Thr Ser Lys Leu Phe Pro Lys Met Pro Gly Leu		
2000	2005	2010
Pro Phe Ile Ser Cys Gln Lys Gly Tyr Lys Gly Val Trp Ala Gly		
2015	2020	2025
Thr Gly Ile Met Thr Thr Arg Cys Pro Cys Gly Ala Asn Ile Ser		
2030	2035	2040
Gly Asn Val Arg Leu Gly Ser Met Arg Ile Thr Gly Pro Lys Thr		
2045	2050	2055
Cys Met Asn Ile Trp Gln Gly Thr Phe Pro Ile Asn Cys Tyr Thr		
2060	2065	2070
Glu Gly Gln Cys Val Pro Lys Pro Ala Pro Asn Phe Lys Ile Ala		
2075	2080	2085
Ile Trp Arg Val Ala Ala Ser Glu Tyr Ala Glu Val Thr Gln His		
2090	2095	2100
Gly Ser Tyr His Tyr Ile Thr Gly Leu Thr Thr Asp Asn Leu Lys		
2105	2110	2115

Val Pro Cys Gln Leu Pro Ser Pro Glu Phe Phe Ser Trp Val Asp		
2120	2125	2130
Gly Val Gln Ile His Arg Phe Ala Pro Ile Pro Lys Pro Phe Phe		
2135	2140	2145
Arg Asp Glu Val Ser Phe Cys Val Gly Leu Asn Ser Phe Val Val		
2150	2155	2160
Gly Ser Gln Leu Pro Cys Asp Pro Glu Pro Asp Thr Asp Val Leu		
2165	2170	2175
Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu Thr Ala		
2180	2185	2190
Ala Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Glu Ala Ser Ser		
2195	2200	2205
Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Arg Ala Thr Cys Thr		
2210	2215	2220
Thr His Gly Lys Ala Tyr Asp Val Asp Met Val Asp Ala Asn Leu		
2225	2230	2235
Phe Met Gly Gly Asp Val Thr Arg Ile Glu Ser Glu Ser Lys Val		
2240	2245	2250
Val Val Leu Asp Ser Leu Asp Pro Met Val Glu Glu Arg Ser Asp		
2255	2260	2265
Leu Glu Pro Ser Ile Pro Ser Glu Tyr Met Leu Pro Lys Lys Arg		
2270	2275	2280
Phe Pro Pro Ala Leu Pro Ala Trp Ala Arg Pro Asp Tyr Asn Pro		
2285	2290	2295
Pro Leu Val Glu Ser Trp Lys Arg Pro Asp Tyr Gln Pro Ala Thr		
2300	2305	2310
Val Ala Gly Cys Ala Leu Pro Pro Pro Lys Lys Thr Pro Thr Pro		
2315	2320	2325
Pro Pro Arg Arg Arg Arg Thr Val Gly Leu Ser Glu Ser Ser Ile		

2330	2335	2340
Ala Asp Ala Leu Gln Gln Leu Ala Ile Lys Ser Phe Gly Gln Pro		
2345	2350	2355
Pro Pro Ser Gly Asp Ser Gly Leu Ser Thr Gly Ala Asp Ala Ala		
2360	2365	2370
Asp Ser Gly Ser Arg Thr Pro Pro Asp Glu Leu Ala Leu Ser Glu		
2375	2380	2385
Thr Gly Ser Ile Ser Ser Met Pro Pro Leu Glu Gly Glu Pro Gly		
2390	2395	2400
Asp Pro Asp Leu Glu Pro Glu Gln Val Glu Leu Gln Pro Pro Pro		
2405	2410	2415
Gln Gly Gly Val Val Thr Pro Gly Ser Gly Ser Gly Ser Trp Ser		
2420	2425	2430
Thr Cys Ser Glu Glu Asp Asp Ser Val Val Cys Cys Ser Met Ser		
2435	2440	2445
Tyr Ser Trp Thr Gly Ala Leu Ile Thr Pro Cys Ser Pro Glu Glu		
2450	2455	2460
Glu Lys Leu Pro Ile Asn Pro Leu Ser Asn Ser Leu Leu Arg Tyr		
2465	2470	2475
His Asn Lys Val Tyr Cys Thr Thr Ser Lys Ser Ala Ser Leu Arg		
2480	2485	2490
Ala Lys Lys Val Thr Phe Asp Arg Met Gln Ala Leu Asp Ala His		
2495	2500	2505
Tyr Asp Ser Val Leu Lys Asp Ile Lys Leu Ala Ala Ser Lys Val		
2510	2515	2520
Thr Ala Arg Leu Leu Thr Leu Glu Glu Ala Cys Gln Leu Thr Pro		
2525	2530	2535
Pro His Ser Ala Arg Ser Lys Tyr Gly Phe Gly Ala Lys Glu Val		

2540	2545	2550
Arg Ser Leu Ser Gly Arg Ala Val Asn His Ile Lys Ser Val Trp		
2555	2560	2565
Lys Asp Leu Leu Glu Asp Thr Gln Thr Pro Ile Pro Thr Thr Ile		
2570	2575	2580
Met Ala Lys Asn Glu Val Phe Cys Val Asp Pro Thr Lys Gly Gly		
2585	2590	2595
Lys Lys Ala Ala Arg Leu Ile Val Tyr Pro Asp Leu Gly Val Arg		
2600	2605	2610
Val Cys Glu Lys Met Ala Leu Tyr Asp Ile Thr Gln Lys Leu Pro		
2615	2620	2625
Gln Ala Val Met Gly Ala Ser Tyr Gly Phe Gln Tyr Ser Pro Ala		
2630	2635	2640
Gln Arg Val Glu Phe Leu Leu Lys Ala Trp Ala Glu Lys Lys Asp		
2645	2650	2655
Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val		
2660	2665	2670
Thr Glu Arg Asp Ile Arg Thr Glu Glu Ser Ile Tyr Arg Ala Cys		
2675	2680	2685
Ser Leu Pro Glu Glu Ala His Thr Ala Ile His Ser Leu Thr Glu		
2690	2695	2700
Arg Leu Tyr Val Gly Gly Pro Met Phe Asn Ser Lys Gly Gln Thr		
2705	2710	2715
Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser		
2720	2725	2730
Met Gly Asn Thr Ile Thr Cys Tyr Val Lys Ala Leu Ala Ala Cys		
2735	2740	2745
Lys Ala Ala Gly Ile Ile Ala Pro Thr Met Leu Val Cys Gly Asp		

2750	2755	2760
Asp Leu Val Val Ile Ser Glu Ser Gln Gly Thr Glu Glu Asp Glu		
2765	2770	2775
Arg Asn Leu Arg Ala Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala		
2780	2785	2790
Pro Pro Gly Asp Pro Pro Arg Pro Glu Tyr Asp Leu Glu Leu Ile		
2795	2800	2805
Thr Ser Cys Ser Ser Asn Val Ser Val Ala Leu Gly Pro Gln Gly		
2810	2815	2820
Arg Arg Arg Tyr Tyr Leu Thr Arg Asp Pro Thr Thr Pro Ile Ala		
2825	2830	2835
Arg Ala Ala Trp Glu Thr Val Arg His Ser Pro Val Asn Ser Trp		
2840	2845	2850
Leu Gly Asn Ile Ile Gln Tyr Ala Pro Thr Ile Trp Ala Arg Met		
2855	2860	2865
Val Leu Met Thr His Phe Phe Ser Ile Leu Met Ala Gln Asp Thr		
2870	2875	2880
Leu Asp Gln Asn Leu Asn Phe Glu Met Tyr Gly Ala Val Tyr Ser		
2885	2890	2895
Val Ser Pro Leu Asp Leu Pro Ala Ile Ile Glu Arg Leu His Gly		
2900	2905	2910
Leu Asp Ala Phe Ser Leu His Thr Tyr Thr Pro His Glu Leu Thr		
2915	2920	2925
Arg Val Ala Ser Ala Leu Arg Lys Leu Gly Ala Pro Pro Leu Arg		
2930	2935	2940
Ala Trp Lys Ser Arg Ala Arg Ala Val Arg Ala Ser Leu Ile Ser		
2945	2950	2955
Arg Gly Gly Arg Ala Ala Val Cys Gly Arg Tyr Leu Phe Asn Trp		

2960	2965	2970
Ala Val Lys Thr Lys Leu Lys Leu Thr Pro Leu Pro Glu Ala Arg		
2975	2980	2985
Leu Leu Asp Leu Ser Ser Trp Phe Thr Val Gly Ala Gly Gly Gly		
2990	2995	3000
Asp Ile Tyr His Ser Val Ser Arg Ala Arg Pro Arg Leu Leu Leu		
3005	3010	3015
Leu Gly Leu Leu Leu Leu Phe Val Gly Val Gly Leu Phe Leu Leu		
3020	3025	3030
Pro Ala Arg		
3033		

【手続補正2】

【補正対象書類名】明細書

【補正対象項目名】図面の簡単な説明

【補正方法】変更

【補正内容】

【図面の簡単な説明】

【図1】は、本発明のNANB肝炎ウイルス遺伝子のコード領域の構造。

【図2】は、NANB肝炎ウイルスの5'末端側塩基配列決定の方法を示す図。HC-J1株。

【図3】は、NANB肝炎ウイルスの5'末端側塩基配列決定の方法を示す図。HC-J4株。

20 【図4】は、NANB肝炎ウイルスの5'末端側塩基配列決定の方法を示す図。HC-J6株。

【図5】HC-J6株ゲノムの3'末端側塩基配列の決定方法。

【符号の説明】

図1において、Cはコア、Eはエンベロープ、NS-1は非構造蛋白質-1、NS-2は同-2、NS-3は同-3、NS-4は同-4、NS-5は同-5。

フロントページの続き

(51) Int. Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
C 0 7 K 13/00		8517-4H		
C 1 2 P 21/02	C	8214-4B		
21/08		8214-4B		
G 0 1 N 33/53	D	8310-2J		
33/576	Z	9015-2J		
33/577	B	9015-2J		
// A 6 1 B 10/00	M			